

\*\*\*\*\*  
M P S R L H  
(TM)  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
638.057 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Perfect Score: 3721  
Sequence: 1 MALWRPSDNTVLPSPVAR.....APSATTSSKPAKRVVRARK 507

Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev

Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query Match %		Description		Pred. No.	
Result No.	Score	Match	Length	DB	ID	Description		Pred. No.	
1	3704	99.5	568	3	PIWL18	L1 protein - human p	0.00e+00		
2	3398	91.3	539	6	S36566	late protein - human	0.00e+00		
3	3052	82.0	505	3	PIWLP3	L1 protein - human p	0.00e+00		
4	3030	81.4	503	3	PIWL39	L1 protein - human p	0.00e+00		
5	2751	73.9	531	6	S36537	late protein - human	0.00e+00		
6	2723	73.2	522	6	S36554	late protein - human	0.00e+00		
7	2690	72.3	503	6	S36549	late protein - human	0.00e+00		
8	2684	72.1	508	6	S36508	late protein - human	0.00e+00		
9	2656	71.4	499	6	S36531	late protein - human	0.00e+00		
10	2644	71.1	534	6	S36583	late protein - human	0.00e+00		
11	2643	71.0	504	3	PIWL51	L1 protein - human p	0.00e+00		
12	2624	70.5	504	3	PIWL31	L1 protein - human p	0.00e+00		
13	2600	69.9	502	14	S36526	late protein - human	0.00e+00		
14	2592	69.7	510	3	S15620	L1 protein - human p	0.00e+00		
15	2589	69.6	531	3	PIWLHS	L1 protein - human p	0.00e+00		
16	2567	69.0	510	3	S15627	L1 protein - human p	0.00e+00		
17	2567	69.0	504	6	S36502	late protein - human	0.00e+00		
18	2557	68.7	505	3	PIWL35	L1 protein - human p	0.00e+00		
19	2549	68.5	524	3	PIWL58	L1 protein - human p	0.00e+00		
20	2550	68.5	528	6	S36520	late protein - human	0.00e+00		
21	2546	68.4	500	3	PIWL6	L1 protein - human p	0.00e+00		

22	2547	68.4	505	6	S36589	late protein - human	0.00e+00
23	2540	68.3	499	3	PIWL33	L1 protein - human p	0.00e+00
24	2542	68.3	502	3	PIWLC1	L1 protein - pygmy c	0.00e+00
25	2542	68.3	505	6	S36560	late protein - human	0.00e+00
26	2526	67.9	529	6	S36578	late protein - human	0.00e+00
27	2520	67.7	499	3	PIWL13	L1 protein - human p	0.00e+00
28	2512	67.5	501	3	PIWL11	L1 protein - human p	0.00e+00
29	2484	66.8	503	6	S36514	late protein - human	0.00e+00
30	2475	66.5	502	3	PIWL42	L1 protein - human p	0.00e+00
31	1988	53.4	508	3	PIWL	L1 protein - human p	0.00e+00
32	1981	53.2	546	6	S36490	late protein - human	0.00e+00
33	1966	52.8	517	6	S36496	late protein - human	0.00e+00
34	1960	52.7	509	6	S36572	late protein - human	0.00e+00
35	1962	52.7	518	14	S36472	late protein - human	0.00e+00
36	1950	52.4	505	3	PIWLRB	L1 protein - cottont	0.00e+00
37	1949	52.4	507	6	S36484	late protein - human	0.00e+00
38	1945	52.3	514	3	PIWL8	L1 protein - human p	0.00e+00
39	1938	52.1	517	6	S36543	late protein - human	0.00e+00
40	1932	51.9	507	6	S36478	late protein - human	0.00e+00
41	1923	51.7	507	6	S36595	late protein - human	0.00e+00
42	1917	51.5	514	3	PIWL47	L1 protein - human p	0.00e+00
43	1908	51.3	501	3	PIWLB2	L1 protein - bovine	0.00e+00
44	1906	51.2	501	3	PIWLEP	L1 protein - Europea	0.00e+00
45	1907	51.2	516	3	PIWL5	L1 protein - human p	0.00e+00

ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE L1 protein - human papillomavirus type 18  
ORGANISM #normal\_name human papillomavirus type 18  
DATE 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 24-Feb-1994  
ACCESSIONS A26251  
REFERENCE A92937  
#authors Cole, S.T.; Danos, O.  
#journal J. Mol. Biol. (1987) 193:599-608  
#title Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products.  
#cross-references MUID:87283882  
#accession A26251  
#molecule\_type DNA  
#residues 1-568 #label COL  
CLASSIFICATION #superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 568 #molecular-weight 63741 #checksum 6794

Query Match				99.5%; Score 3704; DB 3; Length 568;			
Best Local Similarity				99.4%; Pred. No. 0.00e+00;			
Matches				504; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
Db	62	malwrpsdntvlp	ppsvavrvntddytptsifv	hagssrlltvgnpvfrvpaggnkq	121		
QY	1	MALWRPSDNTVLP	PPSVAVRVNTDDYTRTSIFV	HAGSSRLLTGNGPFRVPAGGNKQ	60		
Db	122	dipkvsaygyrvr	qglpdpnkfgldtsiynpetqrlw	wacagveigrqplvgvlgsh	181		
QY	61	DIPKVSAYGYRV	RQGLPDPNKFGLPDNSIYNP	ETQRLWACAGVEIGRQPLVGVL	120		
Db	182	pfynklidtest	haatsnvsvdrnvsvdykqtqclilgc	apalgenwakgtacksrpl	241		
QY	121	PFYNKLIDTEST	HAATSNVSVDRNVSVDYKQTQ	CLILGCAPALGEHWAKGTACKSR	180		
Db	242	sggdcppltel	knvldgmdvtgygandfstlqgtk	cevpldicqscickydyiqmsad	301		
QY	181	SGGDCPPLELKN	TVLDGMDVDTGYGAMDFSLQ	TKEVPLDICQSCICKYDYIQMS	240		
Db	302	pygdsmffcl	rreqifarfhnragtmgdtvpqsl	iykgtgmraspgscvpspsg	361		
QY	241	PYGDSMFFCL	RREQIFARHFNRAAGTMGDT	VPQSLIYKGTGMRASPGSCVPS	360		

```
Db 362 tsdaqlfknpywlhkaqhnngvcwhnqlfvtdtrstnlcastqspvpgqydatk 421
Qy 301 TSDSOLFENKPYWLHKAQHNGNHCWHNQLFVTVVDTTRSTNLTCASQSPVPGQYDATK 360
Db 422 fkyqsrhveeydlqfqlctitladvmsyihmssiledvngfvppttslvdtyr 481
Qy 361 FKQYSRHVEYDLQFQLCTITLADVMSYIHMSSILEDVNGFVPPPTTSLVDTYR 420
Db 482 fvqsvaitcokdaapaenkdpdykllfwnvdlkxfslldoyplgrkflvqaglrkpt 541
Qy 421 FVQSVAITCOKDAAPAEKNDPDKLKFNNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 542 lgrkrsapsattsskprkvrzrk 568
Qy 481 lgrkrsapsattsskprkvrzrk 507

RESULT 2
ENTRY #type complete
TITLE late protein - human papillomavirus type 45
ORGANISM #formal_name human papillomavirus type 45
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36566
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36566
#status preliminary
#molecule_type DNA
#residues 1-539 #label DEL
#cross-references EMBL:X74479
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 539 #molecular-weight 60310 #checksum 9613

Query Match 91.3%; Score 3398; DB 6; Length 539;
Best Local Similarity 87.3%; Pred. No. 0.00e+00;
Matches 448; Conservative 40; Mismatches 19; Indels 6; Gaps 4;

Db 27 malwrpsdntvylpppsvavrvstddvyrtsifvfhagsrlltvgpnyfrvpngagnk 86
Qy 1 MALWRPSDNTVYLPSPSVARVNTDDYVTRTSIFVHAGSRLLTVGNPFRVPAGGNK 59
Db 87 qavkvaayqyrvfvalpdpnkfglpdstlynpetqrlwacvmeigrqplglsq 146
Qy 60 QIPKVSAYQYRVFVQLPDPNKNFGLPDINSYINPETQRLWACAGVEIGRQPLGVLSG 119
Db 147 hpfynkloddteshaataavitqdrndvsvdykqtqlciligcvpaigehwakgtlckpaq 206
Qy 120 HPFYNKLDDESSHAATSNVSEDRDVSVDYKQTQLCILGCAPAIGEHWAKGTACKSRP 179
Db 207 lqpgdcplelknltiedgdmvdtgygamdfstlqdkcevpldicgskypdyiqmsa 266
Qy 180 LSQGDCCPLELKNLTVEDDGDMVDYTGAMDFSTLQDTKCEVPLDICQICKYDYQLQMSA 239
Db 267 dpygdsmfcllrreqifarfhnrgvmgdtvptdilyikgtseannrepqscvyspspg 326
Qy 240 DPYGDSMFCLLRREQIFARHFNWNRAGTNGDTPVQPSLYIKGTG--NRASPGSCVYSPPSG 297
Db 327 slitsdsqfknpywlhkaqhnngicwhnqlfvtdtrstnlcastqnpvpstyd 386
Qy 298 SVTSDSOLFENKPYWLHKAQHNGNHCWHNQLFVTVVDTTRSTNLTCASQSPVPGQYD 357
Db 387 ptkfkqysrhveeydlqfqlctitlaevmsyihmssiledvngfvppttslvd 446
Qy 358 ATKFKQYSRHVEEYDLQFQLCTITLADVMSYIHMSSILEDVNGFVPPPTTSLVD 417
Db 447 tyrfqsvavtcqkdtpekdpydkllkfvtdlkekfasdlqyplgrkflvqaglr 506
Qy 418 TYRFQSVAITCOKDAAPAEKNDPDKLKFNNVDLKEKFSLDLDQYPLGRKFLVQAGLR 477
```

---

```
Db 507 rptigprkraeastastaspkrvrrskk 539
Qy 478 KFTIGPKRRSAPSA-T--TSSKPAKRVRRARK 507

RESULT 3
ENTRY #type complete
TITLE L1 protein - human papillomavirus type ME180 (provirus)
ORGANISM #formal_name human papillomavirus type ME180
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
27-Jan-1995
ACCESSIONS B40509
REFERENCE A40509
#authors Reuter, S.; Delius, H.; Kahn, T.; Hofmann, B.; zur Hausen, H.; Schwarz, E.
#journal J. Virol. (1991) 65:5564-5568
#title Characterization of a novel human papillomavirus DNA in the cervical carcinoma cell line ME180.
#cross-references MUID:91374616
#accession B40509
#molecule_type DNA
#residues 1-505 #label REU
#cross-references GB:M73258
#note translation of the nucleotide sequence is not given
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 505 #molecular-weight 56805 #checksum 4496

Query Match 82.0%; Score 3052; DB 3; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 386; Conservative 76; Mismatches 41; Indels 0; Gaps 0;

Db 1 malwrssdmvylpppsvavrvntddvyrtrtgyvvgagtrlltvgpnyfrvpngagnk 60
Qy 1 MALWRPSDNTVYLPSPSVARVNTDDYVTRTSIFVHAGSRLLTVGNPFRVPAGGNK 60
Db 61 dipkvaayqyrvfvalpdpnkfglpdstlynpetqrlwacvmeigrqplglsq 120
Qy 61 DIPKVSAYQYRVFVQLPDPNKNFGLPDINSYINPETQRLWACAGVEIGRQPLGVLSG 120
Db 121 plynrlidatespfsnknpkardnvadvykqatcligcvpaigehwakgckpsnv 180
Qy 121 PFYNRLDDESSHAATSNVSEDRDVSVDYKQTQLCILGCAPAIGEHWAKGTACKSRPL 180
Db 181 qpgdcpplelknltiedgdmvdtgygamdfstlqdkcevpldicgskypdyiqmsa 240
Qy 181 SQGDCCPLELKNLTVEDDGDMVDYTGAMDFSTLQDTKCEVPLDICQICKYDYQLQMSA 240
Db 241 vvgdsmfcllrreqifarfhnrgvmgdtvptdilyikgtidrdspssyypspsgmv 300
Qy 241 PYGDSMFCLLRREQIFARHFNWNRAGTNGDTPVQPSLYIKGTGMRASPGSCVYSPPSGSIV 300
Db 301 sdsqfknpywlhkaqhnngicwhnqlfvtdtrstnlsttstasvnpnydpnk 360
Qy 301 TSDSOLFENKPYWLHKAQHNGNHCWHNQLFVTVVDTTRSTNLTCASQSPVPGQYDATK 360
Db 361 fkeyrhrveeydlqfqlctitladvmsyihmssiledvngfvppttslvdtyr 420
Qy 361 FKQYSRHVEYDLQFQLCTITLADVMSYIHMSSILEDVNGFVPPPTTSLVDTYR 420
Db 421 ylgasaitcokdaapekdkpdykllfwnvnlkekfsfseidqfplgrkflvqaglrkpt 480
Qy 421 FVQSVAITCOKDAAPAEKNDPDKLKFNNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 481 lgrkrsapsattsskprkvr 503
Qy 481 lgrkrsapsattsskprkvr 503

RESULT 4
ENTRY #type complete
TITLE L1 protein - human papillomavirus type 39
```

```

ORGANISM #formal_name human papillomavirus type 39
#note host Homo sapiens (man)
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
27-Jan-1995
ACCESSIONS H38502
REFERENCE A38502
#authors Volpers, C.; Streeck, R.E.
#journal Virology (1991) 181:419-423
#title Genome organization and nucleotide sequence of human papillomavirus type 39.
#cross-references MIMD:91135017
#accession H38502
#molecule_type DNA
#residues 1-505 #label VOL
#cross-references EMBL:M38185
#note translation of the nucleic acid sequence is not given
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 505 #molecular-weight 56604 #checksum 2488
Query Match 81.4%; Score 3030; DB 3; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 389; Conservative 76; Mismatches 40; Indels 2; Gaps 2;

Db 1 mamwrsdmsvylpppsvakvntddvtrtgytgyagssrlltvgbpyfkvmnggrkq 60
QY 1 MALWRPNDNTVYLPVPPSVARVNTDDVTRTSIFYHAGSSRLLTGVNPFVPPVAGGNGKQ 60
Db 61 dipkvsayqrvfrvlpdpnkfsipdaslynpetqrlvavcgvvegrgqplvglsgh 120
QY 61 DIPKVSAYQRVFRVQLPDPNKFGLPNSYNPETQLVWACAGVEIGRGQPLVGLSGH 120
Db 121 plynrqdgtens-pfsstnkdnsrdvsvdykqqlcigcypaigehgwkackpnnv 179
QY 121 PFYNKLDDESSHAATNSVSEDVNDVSDYKQQLCILGCAIPAIGHWAKGTACKSRPL 180
Db 180 stgdcpplelvntpdegmdtgygamdfgalqetksevpldicqickypdyqlmsad 239
QY 181 SQGDCPPLELKNVLEDDGMDVDTGYGAMDFSTLQDTRCEVPLDICQICKYDPYQLMSAD 240
Db 240 vygdsmffclrrqlfarhfnwngmvgdaipaqlylkgtdiranpgssvycpssgm 299
QY 241 PYGDSMFFCLRRQLFARHFNWNGMVGDTVPQSLYIKGTGMRASPGSCVYSPSPSGIV 300
Db 300 tsdsqfknkpylwhkagghngicwhnqlfvtvdttrstnftstiesisipdygsk 359
QY 301 TSDSQLENKPYLWHKAGHNGICWHNQLFVTVVDTTRSTNLITICASTQSPVPCQYDATK 360
Db 360 fkeytrhveeydlqfqlctvtltdmasyihthmsnllidnwnfapppssalvdtyr 419
QY 361 FKQYSRHVEEYDLQFIFQLCTITLTADVMSYIHSNMSSILEDWNFGVPPPTTSLVDTYR 420
Db 420 ylgaaatcckdapapekkdydglkfwndlrkfeleldqfqlgrkflqarvrirpt 479
QY 421 FVQSVAITCOKDAAPAKNKPDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 480 igprkpaastssss-atkhkrkrvsk 505
QY 481 IGPRKRSAPSATSSSKPAKRVRRARK 507

RESULT 5
ENTRY #type complete
TITLE late protein - human papillomavirus type 10
ORGANISM #formal_name human papillomavirus type 10
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36537
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36537

```

```

#status preliminary
#molecule_type DNA
#residues 1-531 #label DEL
#cross-references EMBL:X74465
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 531 #molecular-weight 59010 #checksum 1781
Query Match 73.9%; Score 2751; DB 6; Length 531;
Best Local Similarity 67.9%; Pred. No. 0.00e+00;
Matches 342; Conservative 103; Mismatches 57; Indels 2; Gaps 2;

Db 29 malwrsdnlvlypbtvpskvlstddvtrtnlyyyagstrlltvgbpyfipksnkv 88
QY 1 MALWRPNDNTVYLPVPPSVARVNTDDVTRTSIFYHAGSSRLLTGVNPFVPPVAGGNGKQ 60
Db 89 dvpkvsayqrvfrvlpdpnkfglpdarynpdaerlyvactgvevgrgqplvglsgh 148
QY 61 DIPKVSAYQRVFRVQLPDPNKFGLPNSYNPETQLVWACAGVEIGRGQPLVGLSGH 120
Db 149 plynkldtensnlahgplgqgdnsldvsnkqqlcigctppmgehgwkgtpcrrpp- 207
QY 121 PFYNKLDDESSHAATNSVSEDVNDVSDYKQQLCILGCAIPAIGHWAKGTACKSRPL 180
Db 208 agdcppleltspigdgmndtgygamdfaltqlnksdvpldicqickypdyqlmaae 267
QY 181 SQGDCPPLELKNVLEDDGMDVDTGYGAMDFSTLQDTRCEVPLDICQICKYDPYQLMSAD 240
Db 268 pygdsmffclrrqlfarhfnrasavgdapdtflknsnggrdvgsavysptpsgm 327
QY 241 PYGDSMFFCLRRQLFARHFNWNGMVGDTVPQSLYIKGTGMRASPGSCVYSPSPSGIV 300
Db 328 tseaqfknkpylwhkagghngicwhnqlfvtvdttrstnftstiesisipdygsk 387
QY 301 TSDSQLENKPYLWHKAGHNGICWHNQLFVTVVDTTRSTNLITICASTQSPVPCQYDATK 360
Db 388 fkeytrhveeydlqfqlctvtltdmasyihthmsnllidnwnfapppssalvdtyr 447
QY 361 FKQYSRHVEEYDLQFIFQLCTITLTADVMSYIHSNMSSILEDWNFGVPPPTTSLVDTYR 420
Db 448 flssaaitcckdapapekkdydglkfwndlrkfeleldqfqlgrkflqarvrirsa 507
QY 421 FVQSVAITCOKDAAPAKNKPDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 508 vsvrkripaastssss-atkhkrkrvsk 530
QY 481 IGPRKRSAPSATSSSKPAKRVRRAR 504

RESULT 6
ENTRY #type complete
TITLE late protein - human papillomavirus type 3
ORGANISM #formal_name human papillomavirus type 3
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36554
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36554
#status preliminary
#molecule_type DNA
#residues 1-532 #label DEL
#cross-references EMBL:X74462
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 532 #molecular-weight 59194 #checksum 1863
Query Match 73.2%; Score 2723; DB 6; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

Db 29 malwrsdnlvlypbtvpskvlstddvtrtnlyyyagstrlltvgbpyfipksnkv 88

```

```
Qy 1 MALWRPSDNTVYLPVPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPYFRVPAGGKNK 60
Db 89 dlpkysaayvrvfvpdpnkfipdpdriynpdaerlvwactvevargipglvqslgh 148
Qy 61 DIPKVSAYQYRVRVQLPDPNKEGLPDNSIYNPETQRLVWACAGVEIGRGQPLGVGLSGH 120
Db 149 plynkliddenslahgdikgkdrdnisvdkqtkqlcivcctpgehwgktckqna- 207
Qy 121 PFYNKLDDESSHAATSNVSEYDRDNVSDYKQQLCILCAPAIGEHWAGTACKSRPL 180
Db 208 spgdcppellitapigdmvdygandfnglsgnsksdvpidicctckypdyylmaae 267
Qy 181 SQGDCPPLELKNVTLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICQICKPTDYLQMSAD 240
Db 268 pygdsmfylrkeqlfarhflnragmagdtvpydalykgsdgsggrdkgsavycptpsg 327
Qy 241 PYGDSMFCLRRREQLFARHFNWNRAGTMDGVTPQSLYIKG---TGMRASPGSCVYSPSPSG 297
Db 328 smvtsetqlfnkpylrragghnngicwancglfvvtvdttrstnmtlcvtstetse-yyd 385
Qy 298 SIYVSDSOLFKNKPYLWHLKAQHNGNICHWNQLFVTVVDTTRSTNLTICASTQSPVPGQYD 357
Db 386 atkfkelylrhgeeydlqfqlckvcltpelmaylhtmnstlledwnfgltlpstaled 445
Qy 358 ATKFKQYSRHVEEYDLQFIFQLCTITLADVMSYIHSNNSIILEDWNFGVPPPTTSLVD 417
Db 446 tyrfitsaitcqkdapptekqpyakinfvdvldkrfaldlsqfgrkfrlmqlgvgt 505
Qy 418 TYRFVQSVAITCQKDAAPAEKNKDPDKLKFNVDLKEKFSLDLDQYPLGRKFLVQAGLR 477
Db 506 rasiavkrksa-tttsrtaakrktk 531
Qy 478 KPTIGPKRSAPSATTSKPKAKRVR 504

RESULT 7
ENTRY S36549 #type complete
TITLE late protein - human papillomavirus type 26
ORGANISM #formal_name human papillomavirus type 26
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995

ACCESSIONS S36549
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36549
#status preliminary
#molecule_type DNA
#residues 1-503 #label DEL
#cross-references EMBL:X74472
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 503 #molecular-weight 56327 #checksum 9874

Query Match 72.3% Score 2690; DB 6; Length 503;
Best Local Similarity 67.3%; Pred. No. 0.00e+00;
Matches 337; Conservative 93; Mismatches 68; Indels 3; Gaps 3;

Db 1 malwrpsdntvylpvpdpnkfipdpdriynpdaerlvwactvevargipglvqslgh 59
Qy 1 MALWRPSDNTVYLPVPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPYFRVPAGGKNK 60
Db 60 eipkysayqyrvfvpdpnkfipdpdriynpdaerlvwactvevargipglvqslgh 119
Qy 61 DIPKVSAYQYRVRVQLPDPNKEGLPDNSIYNPETQRLVWACAGVEIGRGQPLGVGLSGH 120
Db 120 plfnkliddenslahatvnaetdrnrvsvdnkqtkqlcigctpplgehwgltckatqt 179
Qy 121 PFYNKLDDESSHAATSNVSEYDRDNVSDYKQQLCILCAPAIGEHWAGTACKSRPL 180
Db 180 qrdcpellissltdgdmldgframdftalqatksdvpidisctckypdyylkmsad 239
Qy 181 SQGDCPPLELKNVTLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICQICKYPTDYLQMSAD 240

Qy 240 tygnsmfflirteqlfarhfnkagavgdaipttlylkgaesgrepttsiysatpsgsm 299
Qy 241 PYGDSMFCLRRREQLFARHFNWNRAGTMDGVTPQSLYIKGTM-RASPGSCVYSPSPSGSI 299
Db 300 vtadacqlfnkpylrragghnngicwancglfvctvdttrstnlticastaastpfkps 359
Qy 300 VTSDSOLFKNKPYLWHLKAQHNGNICHWNQLFVTVVDTTRSTNLTICASTQSPVPGQYDAT 359
Db 360 dykqfirhgeeyelqfqlckltittdmaylhtmnstlledwnfgltlpptasleday 419
Qy 360 KFKQYSRHVEEYDLQFIFQLCTITLADVMSYIHSNNSIILEDWNFGVPPPTTSLVDTY 419
Db 420 rfnksattcqnnapvvpkedpfqkfkfvdvldkfkfaldldqfplgrkfrlmqlgvgt 479
Qy 420 RFVQSVAITCQKDAAPAEKNKDPDKLKFNVDLKEKFSLDLDQYPLGRKFLVQAGLRKP 479
Db 480 klgt-xrplsstssstkrkr 499
Qy 480 TIGPKRSAPSATTSKPKAKR 500

RESULT 8
ENTRY S36508 #type complete
TITLE late protein - human papillomavirus type 30
ORGANISM #formal_name human papillomavirus type 30
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995

ACCESSIONS S36508
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36508
#status preliminary
#molecule_type DNA
#residues 1-508 #label DEL
#cross-references EMBL:X74474
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 508 #molecular-weight 56667 #checksum 4931

Query Match 72.1% Score 2684; DB 6; Length 508;
Best Local Similarity 67.1%; Pred. No. 0.00e+00;
Matches 337; Conservative 96; Mismatches 67; Indels 2; Gaps 2;

Db 8 mavrpsaetkylpvtvskvvyvtdayvkrtnlfyhagssrlilavghpyysiskagntk 67
Qy 1 MALWRPSDNTVYLPVPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPYFRVPAGGKNK 60
Db 68 dvpkysayqyrvfvpdpnkfipdpdriynpdaerlvwactvevargipglvqslgh 127
Qy 61 DIPKVSAYQYRVRVQLPDPNKEGLPDNSIYNPETQRLVWACAGVEIGRGQPLGVGLSGH 120
Db 128 plfnkliddensatnqdaesdrdnisvdpkqtkqlcigctpplgehwagktacrsapp 187
Qy 121 PFYNKLDDESSHAATSNVSEYDRDNVSDYKQQLCILCAPAIGEHWAGTACKSRPL 180
Db 188 agqdcpplelknvtledgmdvdtgygandfslqdtkcevppldicqickypdyylkmsad 247
Qy 181 SQGDCPPLELKNVTLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICQICKYPTDYLQMSAD 240
Db 248 aygdsmfylrteqlfarhfnragalgeclpstylykgttnrdpppsvvyvatpsgsmv 307
Qy 241 PYGDSMFCLRRREQLFARHFNWNRAGTMDGVTPQSLYIKGTM-RASPGSCVYSPSPSGSI 300
Db 308 tseqlfnkpylrragghnngicwancglfvctvdttrstnmticattt-let-ynsq 365
Qy 301 TSDSOLFKNKPYLWHLKAQHNGNICHWNQLFVTVVDTTRSTNLTICASTQSPVPGQYDATK 360
Db 366 ikgyrvhveeyelqfqlcklisslaetmaylhtmnstlledwnfgltisppaatsledkr 425
Qy 361 FKQYSRHVEEYDLQFIFQLCTITLADVMSYIHSNNSIILEDWNFGVPPPTTSLVDTYR 420
```



```

Db 426 yvksaitcqdqppaekedplakylkfwvnlqdsfadldqfplgrkflmqglvrtkps 485
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 FVQSVAITCQKDAAPAEKNDPYDKLKFVNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 486 tttkrsaspsstpsakrr 507
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 IGPKRRSAPSATTSSKPAKRVR 502

RESULT 9
ENTRY S36531 #type complete
TITLE late protein - human papillomavirus type 53
ORGANISM #formal_name human papillomavirus type 53
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36531
REFERENCE S36469
#authors Dellus, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36531
#status preliminary
#molecule_type DNA
#residues 1-499 #label DEL
#cross-references EMBL:X74482
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 499 #molecular-weight 55722 #checksum 1339

Query Match 71.4%; Score 2656; DB 6; Length 499;
Best Local Similarity 66.3%; Pred. No. 0.00e+00;
Matches 333; Conservative 98; Mismatches 67; Indels 4; Gaps 3;

Db 1 mavrpsdkvlylptpvsyvittdavvkrtrtiffhagssrlltvgbpyypisksg--ka 58
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MALRPSDNTVYLPPPSVARVNTDDYVTRISIFHAGSSRLLVGNPFRVPAGGNGKQ 60
Db 59 dipkvsafqyrvfrirldpnpkfplpntnifnpdgerlhwacvgleigrgqplgvsgsh 118
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 DIPKVSAYQYRVFRVQLPDPNKFGLPDNSIYNPQRLVWACAGVEIGRGQPLGVLSGH 120
Db 119 plfrldtessiaiqdtapdsrdnsvdqkqtcqlclicgacpaigehwtgtsacrstpt 178
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLQILGCAPAIGEHWAAGTACKSRPL 180
Db 179 tagdcplleinspleddgmdvtgfainfkalgsksdvpldivqstckypdykmsad 238
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 SQGDCPPELEANTVLEDDGMDVDTGIGAMDFSTLQDTKEVPLDICQSIKCKIPDYLQWSAD 240
Db 239 aygdsmyfllrreglfrffragvigeelipndlylkgsgngrdppssvvyatpssgm 298
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 PYGDSMFCLRREQLFARHFNWAGTMDTVPSLYIKGTGMRASPGSCVTSPPSGSIV 300
Db 299 tseaqfknkpylragqhgngicwnqdlftvvtvdttrntamtisattqs--mst-yskq 356
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TSDSOLFKNKPYLHKAQHNNGICWHNQLFVTVVDTTSTNLITCASTQSPVPGQYDATK 360
Db 357 ikqvrvhaeeyelqfvgclckisaevmaylhtnmstlledwniglsppvatsledkrr 416
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 FKQSRHVEEYDLQFIFQLCTITITADVMSYIHSNNSSILEDWNFGVPPPTTSLVDIYR 420
Db 417 yvksaatcqdqpppekdqplskylkfwvnlqnsfadldqfplgrkflmqgvrtkpp 476
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 FVQSVAITCQKDAAPAEKNDPYDKLKFVNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 477 vsskrsaspsstpsakrr 498
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 IGPKRRSAPSATTSSKPAKRVR 502

RESULT 10
ENTRY S36583 #type complete
TITLE late protein - human papillomavirus type 56
ORGANISM #formal_name human papillomavirus type 56

```

```

DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36583
REFERENCE S36469
#authors Dellus, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36583
#status preliminary
#molecule_type DNA
#residues 1-534 #label DEL
#cross-references EMBL:X74483
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 534 #molecular-weight 60160 #checksum 1336

Query Match 71.1%; Score 2644; DB 6; Length 534;
Best Local Similarity 67.0%; Pred. No. 0.00e+00;
Matches 337; Conservative 91; Mismatches 69; Indels 6; Gaps 6;

Db 36 matvrpsenkvyllpvtpvskvvatdsyvkrtisfyhagssrllavghpyysv-tkdntkt 94
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MALRPSDNTVYLPPPSVARVNTDDYVTRISIFHAGSSRLLVGNPFRVPAGGNGKQ 60
Db 95 nlpkvsayqyrvfrirldpnpkfplpntnifnpdgerlhwacvgleigrgqplgaglsgh 154
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 DIPKVSAYQYRVFRVQLPDPNKFGLPDNSIYNPQRLVWACAGVEIGRGQPLGVLSGH 120
Db 155 plfrldtesslannnnviedsrdnsldvqkqtcqlclicgacpaigehwtgavckstqv 214
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLQILGCAPAIGEHWAAGTACKSRPL 180
Db 215 ttgdcplleinspleddgmdvtgfainfkalgsksdvpldivqstckypdykmsad 274
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 SQGDCPPELEANTVLEDDGMDVDTGIGAMDFSTLQDTKEVPLDICQSIKCKIPDYLQWSAD 240
Db 275 aygdsmyfllrreglfrffragvigeelipndlylkgsgngrdppssvvyatpssgm 334
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 PYGDSMFCLRREQLFARHFNWAGTMDTVPSLYIKGTGMRASPGSCVTSPPSGSIV 300
Db 335 tseaqfknkpylragqhgngicwnqdlftvvtvdttrntamtisattqs--qls-kydark 392
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TSDSOLFKNKPYLHKAQHNNGICWHNQLFVTVVDTTSTNLITCASTQSPVPGQYDATK 360
Db 393 inqvlrvhaeeyelqfvgclckisaevmaylhtnmstlledwniglsppvatsledkrr 452
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 FKQSRHVEEYDLQFIFQLCTITITADVMSYIHSNNSSILEDWNFGVPPPTTSLVDIYR 420
Db 453 yvrstaitcqdqpppekdqplakylkfwvnlqdsfadldqfplgrkflmqgltrskpa 512
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 FVQSVAITCQKDAAPAEKNDPYDKLKFVNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 513 vatskrsaspsstst-st-pakrrkr 533
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 IGPKRRSAPSATTSSKPAKRVR 502

RESULT 11
ENTRY P1WLJ51 #type complete
TITLE L1 protein - human papillomavirus type 51
ORGANISM #formal_name human papillomavirus type 51
#note host Homo sapiens (man)
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
27-Jan-1995
ACCESSIONS G40415
REFERENCE A40415
#authors Lungu, O.; Crum, C.P.; Silverstein, S.J.
#journal J. Virol. (1991) 65:4216-4225.
#title Biologic properties and nucleotide sequence analysis of human papillomavirus type 51.
#cross-references MUID:91303675
#accession G40415
#molecule_type DNA
#residues 1-504 #label LUN

```

##cross-references GB:M62877  
##note translation of the nucleotide sequence is not given  
CLASSIFICATION superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 504 #molecular-weight 56314 #checksum 7345

Query Match 71.0%; Score 2643; DB 3; Length 504;  
Best Local Similarity 64.0%; Pred. No. 0.00e+00;

Matches 323; Conservative 112; Mismatches 67; Indels 3; Gaps 3;

Db 1 malwrtndskvylpppvsrvntteytrtgtyyagssrltlghpyfpipkts-tra 59

Qy 1 MALWRPSDNTVYLPPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVPAGGNKQ 60

Db 60 alpkvaafgyrvfrvqlpdpnkfgldpdpnlyndtdrlvwcgvvegrsgqlvglsgh 119

Qy 61 DIPKVSAYQYRFRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLVGLSGH 120

Db 120 plfnkyddtensariangagdvrdntsdvdkqtqlcligcappigehwgttckntpv 179

Qy 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAPAIGEHWAKTACKSRPL 180

Db 180 ppqdcpllelvasvlgqdmldtfgamfaalqatksdvpldisgsvckypdyikmsad 239

Qy 181 SQGDCCPPELKNLTVEGDMDVDTGYGAMDFSTLQDKCEVPLDQCICKYDPYLQMSAD 240

Db 240 tynsnmfllrreglfarhynkylvgvedpndvyikgsgngrdpliesviysatpsgsm 299

Qy 241 PYGDSMFFCLRREQLFARHFNWAGTMDTVPQSLYIKGTGM-RASPGSCVYSPSPSGSI 299

Db 300 itadsgifnkpywlrhagghnglcwngnlfitecvdttrstnlticstataavspftfps 358

Qy 300 VTSDQLFNKPYWLHKAQGHNGICWNLQFLVTVVDTTRSTNLTCASQSPVPGQYDAT 359

Db 359 nfkqyirhceveqlfqlckltlttemavylhmdptileqwnfgltlppsaaleday 418

Qy 360 KFKQYSRHYEEDLQFIQLCTITLTADVMSYIHSNSSLIEDWNFGVPPPTTSLVDY 419

Db 419 rfnrnaatcqdtpqekdpdlakypkfdvdkerfeldldgfalgrkflllqvarkp 478

Qy 420 RFQSVVAITCQDAAPAKNDPDKLKFVNVLKKEFSLDLQYPLGRKFLVQAGLRKP 479

Db 479 rpglkprpassassssakkrvk 503

Qy 480 TIGPRKRSAPSATTSSKPAKRVR 504

RESULT 12

ENTRY PIWL31 #type complete

TITLE L1 protein - human papillomavirus type 31

ORGANISM #formal\_name human papillomavirus type 31

note host Homo sapiens (man)

DATE 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change

24-Feb-1994

ACCESSIONS G32444

REFERENCE A94398

authors Goldsborough, M.D.; Disilvestre, D.; Temple, G.F.; Lorincz, A.T.

#journal Virology (1989) 171:306-311

#title Nucleotide sequence of human papillomavirus type 31: a

#cross-references MUID:89299478

#accession G32444

#molecule\_type DNA

#residues 1-504 #label GOL

COMMENT The DNA sequence was obtained from GenBank, release 61.0.

CLASSIFICATION superfamily papillomavirus L1 protein

KEYWORDS late protein

SUMMARY #length 504 #molecular-weight 56352 #checksum 7380

Query Match 70.5%; Score 2624; DB 3; Length 504;

Best Local Similarity 65.1%; Pred. No. 0.00e+00;

Matches 329; Conservative 102; Mismatches 71; Indels 3; Gaps 3;

Db 1 mslwrpsatvylpppvsrvstdeytrtnlyyhagsarlltvghpyysipksdnpkk 60

Qy 1 MALWRPSDNTVYLPPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVP-AGGKNK 59

Db 61 lvpdkvasgfyrvfrvqlpdpnkfgldpdpnlyndtdrlvwcgvvegrsgqlvglsgh 120

Qy 60 QDIFKVSAYQYRFRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLVGLSG 119

Db 121 hollnklddtenstnyvaggpndnrecismdykqtqlcligcappigehwkgkspcsnna 180

Qy 120 HPFNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAPAIGEHWAKTACKSRP 179

Db 181 itpdcpllelvasvlgqdmldtfgamfaalqatksdvpldisgsvckypdyikmsa 240

Qy 180 LSQGDCCPPELKNLTVEGDMDVDTGYGAMDFSTLQDKCEVPLDQCICKYDPYLQMSA 239

Db 241 epyydtlfflyrreglfarhfnrsgtvgesvtdlyikgsgstatlanstyftpsgsm 300

Qy 240 DFYGDSEFFCLRREQLFARHFNWAGTMDTVPQSLYIKGTGM-RASPGSCVYSPSPSGSI 299

Db 301 vtsdaqifnkpywlrhagghnglcwngnlfitecvdttrstnlticstataavspftfps 359

Qy 300 VTSDQLFNKPYWLHKAQGHNGICWNLQFLVTVVDTTRSTNLTCASQSPVPGQYDAT 359

Db 360 nfkqyirhceveqlfqlckltlttemavylhmdptileqwnfgltlppsaaleday 419

Qy 360 KFKQYSRHYEEDLQFIQLCTITLTADVMSYIHSNSSLIEDWNFGVPPPTTSLVDY 419

Db 420 rfnrnaatcqdtpqekdpdlakypkfdvdkerfeldldgfalgrkflllqvarkp 479

Qy 420 RFQSVVAITCQDAAPAKNDPDKLKFVNVLKKEFSLDLQYPLGRKFLVQAGLRKP 479

Db 480 kfkagkrapsattt-pakkrkktk 503

Qy 480 TIGPRKRSAPSATTSSKPAKRVR 504

RESULT 13

ENTRY S36526 #type complete

TITLE late protein - human papillomavirus type 35

ORGANISM #formal\_name human papillomavirus type 35

DATE 09-Dec-1993 #sequence\_revision 09-Dec-1993 #text\_change

09-Dec-1993

ACCESSIONS S36526

REFERENCE S36469

authors Delius, H.; Hofmann, B.

#submission submitted to the EMBL Data Library, August 1993

#accession S36526

#status preliminary

#residues 1-502 #label DEL

##cross-references EMBL:X7477

SUMMARY #length 502 #molecular-weight 56148 #checksum 2108

Query Match 59.9%; Score 2600; DB 14; Length 502;

Best Local Similarity 65.1%; Pred. No. 0.00e+00;

Matches 329; Conservative 95; Mismatches 78; Indels 3; Gaps 3;

Db 1 mslwrpsatvylpppvsrvstdeytrtnlyyhagsarllavghpyyaikkqdenki 60

Qy 1 MALWRPSDNTVYLPPPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVPAGGNKQ 60

Db 61 avpkvasgfyrvfrvqlpdpnkfgldpdpnlyndtdrlvwcgvvegrsgqlvglsgh 120

Qy 61 DIPKVSAYQYRFRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLVGLSGH 120

Db 121 pllnklddtenstnyvaggpndnrecismdykqtqlcligcappigehwkgktpcnanq 180

Qy 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAPAIGEHWAKTACKSRPL 180

Db 181 kagecpllelntvlgdgmldtfgamdfllqanksdvpldisgsvckypdyikmsv 240

Qy 181 SQGDCCPPELKNLTVEGDMDVDTGYGAMDFSTLQDKCEVPLDQCICKYDPYLQMSAD 240

Db	241	pygdmllfyylrreqlmrvhrhlnragtvgvtpadilykgttg-tltpats-yfptpsgsmy	298
QY	241	pygdmfmlrreqlmrvhrhlnragtvgvtpadilykgttg-tltpats-yfptpsgsmy	300
Db	299	tsdaqifnkywqlraaghngicwsgnqlfvtvvttrstnmsvcsvassv-dstykndn	357
QY	301	tsdqlfknkpylwhkaghnngicwsgnqlfvtvvttrstnmsvcsvassv-dstykndn	360
Db	358	fkeylrhgeevdqlfqlckititadvmtyihsmnpsilledwnfgitpppsgtiedtyr	417
QY	361	fkeylrhgeevdqlfqlckititadvmtyihsmnpsilledwnfgitpppsgtiedtyr	420
Db	418	yvtsgavtcqpsapkpdkpdknytfwvdlkkekfsadldqfplgrkflqaglkarnp	477
QY	421	fvosvaitcqkdrapasnckpydylkfwvdlkkekfsadldqfplgrkflqaglkarnp	480
Db	478	frlgkraapastkksktrkvtks	502
QY	481	igprksrapsattsskprkvrva	505
RESULT	14		
ENTRY		S15620	#type complete
TITLE		L1 protein - human papillomavirus type 2a	
ORGANISM		#formal_name human papillomavirus type 2a	
#note		host Homo sapiens (man)	
DATE		17-Feb-1994	#sequence_revision 17-Feb-1994 #text_change
ACCESSIONS		S15620	
REFERENCE		S15614	
#authors		Hirsch-Behnam, A.; Dellus, H.; de Villiers, E.M.	
#journal		Virus Res. (1990) 18:81-98	
#title		A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57.	
#cross-references		MUID:91188699	
#accession		S15620	
#molecule_type		DNA	
#residues		1-510	#label HIR
#cross-references		EMBL:X55964	
CLASSIFICATION		#superfamily papillomavirus L1 protein	
KEYWORDS		late protein	
SUMMARY		#length 510 #molecular-weight 57193 #checksum 5538	
Query Match		69.7%; Score 2592; DB 3; Length 510;	
Best Local Similarity		64.1%; Pred. No. 0.00e+00;	
Matches		323; Conservative 111; Mismatches 59; Indels 11; Gaps 9	
Db	17	malwrpneskvyllpdpvskvistdvvtntvvyhbgssrlitvghpyvsikksn-nkv	75
QY	1	malwrpsdntvylpppsvarvntddvtrtsifyhagssrllltvgnfyfvpagggnkq	60
Db	76	avpkvsqyrvfhvklpdpnkgfplpadilydpdqttrllwacvgevgrgqplgvsgsh	135
QY	61	dipkvsayqrvfrvqlpdpnkfplpdsniynpetqrlwacagveigrqplgvglsgl	120
Db	136	pyynrltdtenah--tpdtaddgrenismdykqtqlfclgckppigehwskgttcngss-	192
QY	121	pfynklddtesshaatnsnvdsdvndsvdykqtqlcglcapaigshwagkacksrpl	180
Db	193	aagdcplqlfnttiedgdmvetqfgaldfatigsksdvpldctctkypdykmaae	252
QY	181	sgdcplpqlkntvleogdavidgvgandstlqdtkcevpldicqickypdykmaae	240
Db	253	pygdmfmlrreqlmrvhrhlnragtvgvtpadilykgttg-tltpats-yfptpsgsmy	311
QY	241	pygdmfmlrreqlmrvhrhlnragtvgvtpadilykgttg-tltpats-yfptpsgsmy	300
Db	312	sseqlfnkywqlraaghngicwsgnqlfvtvvttrstnmsvcsvassv-dstykndn	369
QY	301	tsdqlfknkpylwhkaghnngicwsgnqlfvtvvttrstnmsvcsvassv-dstykndn	360
Db	370	fkeylrhgeevdqlfqlckititadvmtyihsmnpsilledwnfgitpppsgtiedtyr	429

Mon Mar 31 09:21:05 1997

US-08-409-122-2.rpr

Page 8

Search completed: Wed Mar 26 08:29:01 1997  
Job time : 51 secs.

\_\_\_\_\_

\*\*\*\*\*

W P S R L L

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:29:18 1997; Maspar time 11.98 Seconds  
435.828 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Perfect Score: 3721  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507

Scoring table: PAM 150  
Gap 11

Searched: 88003 seqs, 10295656 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq25  
1:part1 2:part3 3:part4 4:part5 5:part6 6:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18

Statistics: Mean 35.688; Variance 147.981; scale 0.241

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2610	70.1	505	8	R38807	4.27e-252
2	2595	69.7	505	8	R38808	1.52e-250
3	2547	68.4	532	17	R30316	1.41e-245
4	1992	53.5	508	4	P30315	2.94e-188
5	1862	50.0	495	3	P51099	7.26e-175
6	1178	21.7	215	16	R88267	9.86e-105
7	764	20.5	221	16	R88273	8.54e-63
8	754	20.3	222	16	R88268	8.62e-62
9	742	19.9	220	16	R88269	1.38e-60
10	728	19.6	224	16	R88271	3.50e-59
11	724	19.5	224	16	R88274	8.80e-59
12	724	19.5	225	16	R88270	8.80e-59
13	717	19.3	220	16	R88272	4.42e-58
14	720	19.3	228	16	R88275	2.21e-58
15	466	12.5	88	6	R32057	3.33e-33
16	466	12.5	88	3	R14299	3.33e-33
17	292	7.8	39	1	P32074	1.66e-16
18	259	7.0	51	6	R32062	1.94e-13
19	259	7.0	51	6	R32059	1.94e-13
20	241	6.5	36	6	R32071	8.70e-12
21	226	6.1	36	6	R32065	2.01e-10

22	217	5.8	35	6	R32063	Human IgA reactive HP	1.30e-09
23	211	5.7	46	6	R32070	VLP reactive HPV16 L1	4.48e-09
24	193	5.2	35	6	R32066	VLP reactive HPV16 L1	1.77e-07
25	184	4.9	39	1	P82075	Hman papilloma virus	1.09e-06
26	164	4.4	26	6	R32068	VLP reactive HPV16 L1	5.79e-05
27	161	4.3	20	1	R04980	Papilloma virus type	1.04e-04
28	157	4.2	37	3	R14301	Seroreactive epitope	2.27e-04
29	151	4.1	20	1	R04966	Papilloma virus type	7.24e-04
30	146	3.9	36	6	R32069	VLP reactive HPV16 L1	1.89e-03
31	137	3.7	20	1	R04984	Papilloma virus type	1.04e-02
32	136	3.7	20	1	R04981	Papilloma virus type	1.25e-02
33	137	3.7	21	3	R15589	Immunopeptide derived	1.04e-02
34	134	3.6	20	1	R04965	Papilloma virus type	1.82e-02
35	133	3.6	88	3	R14298	Seroreactive epitope	2.19e-02
36	130	3.5	20	1	R04985	Papilloma virus type	3.82e-02
37	130	3.5	20	1	R04974	Papilloma virus type	3.82e-02
38	129	3.5	20	1	R04970	Papilloma virus type	4.59e-02
39	127	3.4	20	1	R04964	Papilloma virus type	6.54e-02
40	126	3.4	20	1	R04975	Papilloma virus type	7.97e-02
41	124	3.3	20	1	R04990	Papilloma virus type	1.15e-01
42	123	3.3	20	1	R04976	Papilloma virus type	1.38e-01
43	119	3.2	20	1	R04961	Papilloma virus type	2.84e-01
44	118	3.2	22	3	R14302	Seroreactive epitope	3.40e-01
45	115	3.1	20	6	R32058	Human IgG reactive HP	5.81e-01

ALIGNMENTS

RESULT 1	ID	R38807 standard; Protein; 505 AA.
ID	AC	R38807;
DT	DT	21-JAN-1994 (first entry)
DE	DE	BPV1 L1
KW	KW	L1: capsid protein; bovine; human; papillomavirus; baculovirus;
KW	KW	transfer vector; promoter; capsid protein; transformation; vaccine;
KW	KW	neutralising antibody; vertebrate.
OS	OS	Bovine papillomavirus.
PN	PN	US032869-A.
PD	PD	15-JUL-1993.
PF	PF	03-SEP-1992; 941371.
PR	PR	03-SEP-1992; US-941371.
PR	PR	16-MAR-1993; US-032869.
PA	PA	(USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI	PI	Kirnbauser R, Lowy DR, Schiller JT;
DR	DR	WPI: 93-249995/31.
DR	DR	N-PSDB; Q47166.
PT	PT	Recombinant papilloma virus capsid proteins - for vaccines
PT	PT	against papilloma virus and for diagnosis of virus infection
PS	PS	Example 1; Page 32-34; 45pp; English.
CC	CC	The sequences given in R38807-08 represent the L1 capsid proteins from
CC	CC	bovine and human papillomavirus respectively. The DNA encoding these
CC	CC	sequences may be inserted into a baculovirus transfer vector and
CC	CC	operatively expressed by a promoter of the vector, and the capsid
CC	CC	protein produced by transformed cells. These capsid proteins may be
CC	CC	used in vaccines to induce high-titre neutralising antibody response
CC	CC	in vertebrates.
CC	CC	Sequence 505 AA;
QY	QY	Query Match 70.1%; Score 2610; DB 8; Length 505;
QY	QY	Best Local Similarity 65.5%; Pred. No. 4.27e-252;
QY	QY	Matches 328; Conservative 96; Mismatches 75; Indels 2; Gaps 2;
Db	Db	1 mslwlpseavtlyppvsvkvstdevatnlyyhgatrllyavghpypfipkpnanki 60
QY	QY	1 MALWRPSDNTVYLPSPVARVNTDDYVTRTSIFYHAGSSRLTLTVGNPYFRVPAAGGNKQ 60
Db	Db	61 lvpkvsqlyvrfihlpdnpkfgpdtqrlvwacvvgvgrsqplvgvlgsh 120
QY	QY	61 DIPKVSAYQVRFVQLPDPNKGKLPDINSYNPETQLRWACAGVEICRQPLGVLSGH 120
Db	Db	121 pllnklldtenasayaanaagvdmrcismdykqtqlclgckppighwgkspctnav 180
QY	QY	121 PFYNKLDTESSHAATSNVSEDVRDNDVYDKOTQLCILGCAPIGHEWAKGTACKSRPL 180

Db	181	npgdcpllelntvliqgdvnmhvgfgamdfittlqanksevpldictstickypdyikmwse	240
Qy	181	SQGDPEPLSELKNTVLEDDGMVDGYGAMDFSTLQDTKEVPDLCQSTCKYPDYLOMSAD	240
Db	241	pygdsifvyrrtegmvrhlnfragvtgvengpddlylkgsgatanlasyfptpgsmv	300
Qy	241	PYGDSMFPEFLCRREQULFAHWNHNRAGTMGDTPOQSLYKGTGMRA:SPGCVYSPSPSGSIV	300
Db	301	tsdqifknkpywqlragghnngicwgnqifvtvtdtrstnmalcaaisett-ykntn	359
Qy	301	TSDSOLFENKPYWLHRAQGNHNCWNLQEVTVVDTVTSTNTICASTQSPVQGVYATK	360
Db	360	fkeylrhgeeydlqifgckiltltdavmtyihsmstiledwnfglpppggtleddy	419
Qy	361	FKQYSRIVEEYDLQIFOLCTITLTADVMSXITHSNMSILEDWNFGVPPPTTSLVDYTR	420
Db	420	fvtsqiaiaqckhtppapkedplkkytfwevnlkekfadsadlqdgfgrklllqaglakpk	479
Qy	421	FVQSVATJCQDAAPENKDPYDKLFWNVDLKFKFSLDLDQYPLGRFLVQAGLARKEPT	480
Db	480	ftlgkrkatpattsssttakr	500
Qy	481	IG-PKRSAPSATTSSKPAKR	500

RESULT	2	
ID	R38808	standard; Protein; 505 AA.
AC	R38808	AC
AD	R38808	AD
DT	21-JAN-1994	(first entry)
DE	HPV1	LI.
DI	LI;	capsid protein; bovine; human; papillomavirus; baculovirus;
KW	transfer vector;	promoter; capsid protein; transformation; vaccine;
KN	neutralising antibody;	vertebrate.
KS	Human papillomavirus.	
OS	US8032868-A.	
PN	15-JUL-1993.	
PP	941371.	
PR	03-SEP-1992;	US-941371.
PR	03-SEP-1992;	US-941371.
PR	16-MAR-1993;	US-032869.
PR	(US\$S )	US DEPT HEALTH & HUMAN SERVICE.
PI	Kirnbauer R, Lowy DR, Schiller JT;	
PI	WPI: 93-249995/31.	
PT	N-PADB; Q47167.	
PT	Recombinant papilloma virus capsid proteins - for vaccines	
PT	against papilloma virus and for diagnosis of virus infection	
PS	Example 1; Page 35-37; 45pp; English.	
SC	The sequences given in R38807-08 represent the LI capsid proteins from	
CC	bovine and human papillomavirus respectively. The DNA encoding these	
CC	sequences may be inserted into a baculovirus transfer vector and	
CC	operatively expressed by a promoter of the vector, and the capsid	
CC	protein produced by transformed cells. These capsid proteins may be	
CC	used in vaccines to induce high-titre neutralising antibody response	
CC	in vertebrates.	
CC	Sequence	505 AA:

QY	181	SGGDCPLELKNVTLDEGDMVDTGYGAMDFTLQTKCEVPLDLCQSIKCYDPDLQMSAD	240
Db	241	pygds1lffylrrcmfwhrlfractgvencvpadllyksgsetanlasyfptsgamv	300
QY	241	FIGDSMEFLCRLREQULFAHFHNRACTMGDTVPQSLIKGTGNRASPGSCVYSPSPGSGIV	300
Db	301	tsdqifnkpvyllqraghnglcnwgnqlftvvdttrtnmslcaaisett-ykntn	359
QY	301	TSDSQLEFNKPYLWHAQGHNGICWHNQLFTVVDTTRTNATICATSGFSVPQGVDAIKR	360
Db	360	kfeylrheeayldqfqlckiltadvmtyihamnstllednwfnglqpppggtledtyr	419
QY	361	FKQYSRIVEYDQLQIFQLCTITLTADVNSYIHSNMSSILEDMNFGVPPPTTSLVDYIKR	420
Db	420	fv-tqaiacqkthtpapkeddpikkytfwevnlkefsadllqfplgrkfilcaglkakp	478
QY	421	FVQSVATTCOKDAAPA-ENKDPYDKLFKFWNLDLKEFSLDQYPLGRKFLVQAGLRKP	479
Db	479	kftlgkktatpstsststakr 500	
QY	480	TIG-PRKRSAPSATISSPKAKR 500	
RESULT	3		
ID	883016	standard; Protein; 532 AA.	
AC	883016	1996 (first entry)	
DE	Recombinant papilloma virus L1 protein (HPV6b L1 HEXAHIS protein).		
KW	Human papilloma virus; HPV; virus like particles; VLP; L1 protein;		
KW	vaccine; immune response; antibody; reagent; immunoassay; detection;		
KW	infection.		
OS	Human papilloma virus.		
PN	W09531476-A1.		
PD	23-NOV-1995.		
PF	17-MAY-1995; AU0292.		
PR	17-MAY-1994; AG-005667.		
PA	(UYQU) UNIV QUEENSLAND.		
PI	Frazer I, Zhou J;		
DR	WPI; 96-010877/01.		
PT	N-PDSB; T05834.		
PT	Recombinant papilloma virus L1 protein forming extracellular		
PT	multimer(s) - useful in vaccines, as assay reagent and for producing		
PT	antibodies		
PS	Claim 3; Figure 1; 33pp; English.		
CC	The recombinant papilloma virus L1 protein is able to elicit an		
CC	immune response that recognises papilloma virus-like particles		
CC	(VLP's) and to form extracellular multimeric structures consisting		
CC	of several recombinant L1 protein copies. The recombinant L1		
CC	protein may also be used in vaccines and to raise antibodies. It		
CC	is also useful for the detection of human papilloma virus		
CC	antibodies and hence, as a reagent in an immunoassay. The protein		
CC	may induce both a cellular and humoral immune response.		
SQ	Sequence 532 AA;		
Query Match	68.4%;	Score 2547; DB 17; Length 532;	
Best Local Similarity	63.2%;	Pred. No. 1.41e-245;	
Matches	321; Conservative	107; Mismatches 73; Indels 7; Gaps	
Db	31	kdmwrpsdstvypnpnyskvsatdaytrtnifyhasesrllavghpyfsikra--nk 88	
QY	1	MAIWRPSDNTVYLPSPS-VARVYNTDDYVTRISYFHAGSRLLTVGNPYRFPAGGKN 59	
Db	89	ttvvpkvsygyrvkvvlpdpnkalpdsalfdptqtrvwactglvgrgpplgvsg 148	
QY	60	QDTPKVSAYQYRVERQLPDPNPKFGLPDSINPTQSLWACAGVEIGRQPIUGVLSG 119	
Db	149	hpfinkyddvens-gaggpnpgqdnrvnvmkyqtkqlcmvvgcappihgwkgkcntp 207	
QY	120	HPFYNKLDDESSHAATSNVSEEDVRDNVVDYKQTKCILGCAPAGEHWAGACKSRP 179	
Db	208	vqagdcppleltsvlgdgmvdvctfgfamafdqlgnksdvpidicgtctckypdylgmaa 267	

Qy 180 LSQGDCCPPLKNTVLEGGDMVDYGYGAMDFSTLQDTKEVPLDTCQICKYPPYLQMSA 239  
 Db 268 dpygdlrffirkecmfarhnfagevgevpdpdltliiksgnrtsvsslyvntpsgl 327  
 Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSGSI 299  
 Db 328 vsseqlfnkpywlqacghngicwgnqlfvtvvdtrstnmtlcasv-tt-astyns 385  
 Qy 300 VTSQSLFNKPYLHKAQHNGNHCWNLQFLVTVVDTTRSTNLITICASTQSPVPGQYDAT 359  
 Db 386 dykymrhveeydqfifqlcsittisaemaylhtmpnsvledwfnlgspnngtledty 445  
 Qy 360 KFKQYSRHEVEYDQFIFQLCTITLTADVMSYIHSNMSSILEDWNFVGPPTTSLVDY 419  
 Db 446 ryvsgaitcdkptpekekpdpynksfwevnlkekfeseldqyplgrkflilqsgygrs 505  
 Qy 420 RFVQSVAITCQKDAAPENKDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGLRKP 479  
 Db 506 sirtgvkr-pavskasaapkrakrkr 532  
 Qy 480 TIGPKRSAPSATTSSKPAKRVRAR 507

## RESULT 4

ID P30315 standard; Protein; 508 AA.  
 AC P30315;

DT 05-APR-1992 (first entry)  
 DE Sequence encoded by the L1 region of the human papillomavirus (hpv)

DE type Ia genome.

KW Diagnostic reagent; vaccine; medicine; wart; tumour.

OS Human papillomavirus.

PN EP-92456-A.

PD 26-OCT-1983.

PF 01-APR-1983; 901081.

PR 05-APR-1982; FR-005687.

PA (INSP) INST PASTEUR.

FA (DANO) DANOS O.

PI Danos O, Katinka M, Yaniv M;

DR WPI; 83-802979/44.

DR N-PSDB; N30174.

PT DNA fragment coding for Papillomavirus antigenic proteins - and

PT derived immunogen, vaccine and antibody

PS Disclosure; Fig 1; 25pp; French.

CC The inventors claim DNA fragments capable of expressing, in a host, a prod. contg. at least one antigenic determinant of papillomavirus (PV), (see N30170-N30173). Also claimed are immunogens consisting of at least one peptide sequence coded for by the DNA fragments (see N30310-P30313), vaccines contg. the immunogens and antibodies raised from them. The vaccines are useful in human and veterinary medicine and the antibodies are useful as diagnostic reagents. The DNA fragments are most esp. derived from the L1 region of the human PV type Ia.

CC Sequence 508 AA;

Query Match 53.5%; Score 1992; DB 4; Length 508;

Best Local Similarity 49.4%; Pred. No. 2,946-188;

Matches 250; Conservative 138; Mismatches 107; Indels 11; Gaps 8;

Db 7 mavlpaqnfkylppqpitrilstdeytrtnlfyhatserlllvghplfaissn--qtv 64

Qy 1 MALWRPSDNTYLLPPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNFYRVPAGGNKQ 60

Db 65 tipkvspnafvfrfapdnrfafgkaifnpeterlvwlgieigrqplgigtgh 124

Qy 61 DIPKVSAYQYRVRVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120

Db 125 pllnklldaenptnyinhangderqntafadqgtqmfvgctpasgwht--srrcpgeq 183

Qy 121 PFYNKLDDES-SHAATSNVSEDVDRNVSDYKQTLQILCAPAIGEHAKGACKSRP 179

Db 184 vklgdcpvrvmlesviedgmdndfgandfaalqgdkdsdpdvqatckypdyimnh 243

Qy 180 LSQGDCCPPLKNTVLEGGDMVDYGYGAMDFSTLQDTKEVPLDTCQICKYPPYLQMSA 239

Db 244 eayngsmffarreqmytrhffrtgrgvgdkeavpqslyltadaepttlattnvgtps 303  
 Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSGS 296  
 Db 304 gsmvssdqlfnrswlqroqgnngicwgnqlfvtvvdtrstnmtlcasv-tt-astyns 361  
 Qy 297 GSIIVTSQSLFNKPYLHKAQHNGNHCWNLQFLVTVVDTTRSTNLITICASTQSPVPGQY 356  
 Db 362 snaanfdrflrtheefdlfvlqckvkltpenlaylhtmdpnilledwqlsvsqptuple 421  
 Qy 357 DATFKQYSRHEVEYDQFIFQLCTITLTADVMSYIHSNMSSILEDWNFVGPPTTSLV 416  
 Db 422 dqyiflgsslaakcpeappdpqtdpysqykfwevdltermseqldgfpigrkflilqsgm 481  
 Qy 417 DTYRFV-QSVAITCQKDAAPENKDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGL 475  
 Db 482 tqrtatsttkrktvrvstsaakrrk 507  
 Qy 476 -RRKPTIGPKRSAPSATTSSKPAKR 500

## RESULT 5

ID P51099 standard; Protein; 495 AA.

AC P51099;

DT 21-NOV-1991 (first entry)

DE Sequence of bovine papillomavirus capsid protein L1.

KW Vaccine; papillomavirus; viral capsid protein.

OS Bovine papillomavirus.

PN EP-133123-A.

PD 13-FEB-1985.

PF 23-JUL-1984; 401548.

PR 25-JUL-1983; US-517063.

PR 11-JUL-1984; US-629852.

PA (MOLE-) MOLECULAR GENETICS.

PI Pilacinski WP, Glassman DL, Krzyzek RA, Sadowski PL;

DR WPI; 85-039686/07.

DR N-PSDB; N50262.

PT prodn. of protein(s) related to papilloma virus capsid protein(s)

PT - by recombinant DNA techniques for vaccine prodn.

PS Disclosure; Fig 3; 100pp; English.

CC The inventors claim a recombinant DNA vector, comprising: a DNA sequence coding for an immunoreactive and antigenic determinant of a papillomavirus protein wherein the DNA SQ codes for L1 or L2 as depicted in Fig 3. Using the vector, proteins related to viral capsid proteins of papillomavirus can be obtd. by recombinant DNA techniques and used as vaccines.

CC Sequence 495 AA;

Query Match 50.0%; Score 1862; DB 3; Length 495;

Best Local Similarity 48.8%; Pred. No. 7,26e-175;

Matches 245; Conservative 121; Mismatches 124; Indels 12; Gaps 10;

Db 1 malwqgqgk-lylpptpvskvscsetyvrktsifyhaeterlltighpyy--plsigdk- 56

Qy 1 MALWRPSDNTYLLPPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNFYRVPAGGNKQ 60

Db 57 tvpkvsanqyrvfkiqldpnpqfalpdrthmpskervlwavigvsvrgplggtvtgh 116

Qy 61 DIPKVSAYQYRVRVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120

Db 117 ptfalldaenlnrkvttqtddd-rkqtdakqgqlllqctpaegvwtartpcvtdr 175

Qy 121 PFYNKLDDES-SHAATSNVSEDVDRNVSDYKQTLQILCAPAIGEHAKGACKSRP 179

Db 176 lengacppllelknhiedgdmmeigfgaanfkeinaasksdpldigneiclypdykmae 235

Qy 180 LSQGDCCPPLKNTVLEGGDMVDYGYGAMDFSTLQDTKEVPLDTCQICKYPPYLQMSA 239

Db 236 daagnsmffarkeqyvrvhlwtrggsekeapttdfylvknkkgdatlkipsvfhgspgs 295

Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSGS 296

```
Db 296 lvsdngfnrpyrlfragmngiawnllflvgtngtrgtntltisvpsdgtplteyds 355
Qy 299 IVVSQSLFNKPYWHLHKAQHNGICWHNQLEFVWVDTTRSTNLTICASTQSPVPGQYDA 358
Db 356 skfnvyrhmeeyklafilelscveltaetvshlqglmpsvlenveigvqptasiledt 415
Qy 359 TKFKQISRVVEYDQLQFQLCITITADVMSYIHSNMSILEDNMGVPPPTISLVDI 418
Db 416 yryiespatkcasnvipa-kqdyagfkwidlkqklsldldqfgrflagagagc- 473
Qy 419 YRVQSVATCOKDAEAKNDPYDKLKFVNVDLKEKFLDLDQYPLGRKFLVQAGLRK 478
Db 474 stv--rkrri-sqktskspakk 492
Qy 479 PTIGPKRSAPSANTSSPKAR 500

RESULT
ID R88267 standard; Peptide; 215 AA.
AC R88267;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 29; papilloma virus; major;
KW capsid protein; plasmid VS93-1; DSM 9133; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
DR N-PSDB; T03499.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 7; 15pp; German.
CC Peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS93-1 (DSM 9133). The DNA is useful in
CC diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 215 AA;

Query Match 31.7%; Score 1178; DB 16; Length 215;
Best Local Similarity 70.2%; Pred. No. 9,866-105;
Matches 151; Conservative 37; Mismatches 27; Indels 0; Gaps 0;

Db 1 grgqplvgisghpnylnkndtensniahadpsdrdnsvdkqtqlcylcgtppmge 60
Qy 108 GRGQPLVGLSGHPFYFNKLDDESSHAATSNVSEYDRDNVSDYKQTLQILGCAPAIGE 167
Db 61 ywkgtpcarnttbgdcppllemtsyiqgdmvdtgygandftalqfksdvpldicqs 120
Qy 168 HWAKGTACKSRPLSQDCGCPPLKNTVLEGDVMDVDTGYGAMDFSTLQDTKCEVPLDICQS 227
Db 121 lckypdylnmaadpygdmfflrreglfarhfnragvgdkipesylyksgsgrtppg 180
Qy 228 ICKYPDYLQMSADPYGDSMFFCLRRQELFARHFNWNRAGTMGDTVPQSLYIKGTGMASPG 287
Db 181 saiyptpsgmvtseaqifnksywlagagqngn 215
Qy 288 SCVYSPSPSGSIVTSDSQLENKPYWLHKAQHNGN 322

RESULT
ID R88273 standard; Peptide; 221 AA.
AC R88273;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 15; papilloma virus; major;
KW capsid protein; plasmid VS42-1; DSM 9139; diagnosis;

Db 296 lvsdngfnrpyrlfragmngiawnllflvgtngtrgtntltisvpsdgtplteyds 355
Qy 299 IVVSQSLFNKPYWHLHKAQHNGICWHNQLEFVWVDTTRSTNLTICASTQSPVPGQYDA 358
Db 356 skfnvyrhmeeyklafilelscveltaetvshlqglmpsvlenveigvqptasiledt 415
Qy 359 TKFKQISRVVEYDQLQFQLCITITADVMSYIHSNMSILEDNMGVPPPTISLVDI 418
Db 416 yryiespatkcasnvipa-kqdyagfkwidlkqklsldldqfgrflagagagc- 473
Qy 419 YRVQSVATCOKDAEAKNDPYDKLKFVNVDLKEKFLDLDQYPLGRKFLVQAGLRK 478
Db 474 stv--rkrri-sqktskspakk 492
Qy 479 PTIGPKRSAPSANTSSPKAR 500

RESULT
ID R88267 standard; Peptide; 215 AA.
AC R88267;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 29; papilloma virus; major;
KW capsid protein; plasmid VS93-1; DSM 9133; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
DR N-PSDB; T03499.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 7; 15pp; German.
CC Peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS93-1 (DSM 9133). The DNA is useful in
CC diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 215 AA;

Query Match 31.7%; Score 1178; DB 16; Length 215;
Best Local Similarity 70.2%; Pred. No. 9,866-105;
Matches 151; Conservative 37; Mismatches 27; Indels 0; Gaps 0;

Db 1 grgqplvgisghpnylnkndtensniahadpsdrdnsvdkqtqlcylcgtppmge 60
Qy 108 GRGQPLVGLSGHPFYFNKLDDESSHAATSNVSEYDRDNVSDYKQTLQILGCAPAIGE 167
Db 61 ywkgtpcarnttbgdcppllemtsyiqgdmvdtgygandftalqfksdvpldicqs 120
Qy 168 HWAKGTACKSRPLSQDCGCPPLKNTVLEGDVMDVDTGYGAMDFSTLQDTKCEVPLDICQS 227
Db 121 lckypdylnmaadpygdmfflrreglfarhfnragvgdkipesylyksgsgrtppg 180
Qy 228 ICKYPDYLQMSADPYGDSMFFCLRRQELFARHFNWNRAGTMGDTVPQSLYIKGTGMASPG 287
Db 181 saiyptpsgmvtseaqifnksywlagagqngn 215
Qy 288 SCVYSPSPSGSIVTSDSQLENKPYWLHKAQHNGN 322

RESULT
ID R88273 standard; Peptide; 221 AA.
AC R88273;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 15; papilloma virus; major;
KW capsid protein; plasmid VS42-1; DSM 9139; diagnosis;
```

```
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
DR N-PSDB; T03505.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 7; 15pp; German.
CC Peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS42-1 (DSM 9139). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 221 AA;

Query Match 20.5%; Score 764; DB 16; Length 221;
Best Local Similarity 51.6%; Pred. No. 8,546-63;
Matches 114; Conservative 40; Mismatches 61; Indels 6; Gaps 2;

Db 1 grgqplvgisghpnylnkndtensniahadpsdrdnsvdkqtqlcylcgtppmge 60
Qy 108 GRGQPLVGLSGHPFYFNKLDDESSHAATSNVSEYDRDNVSDYKQTLQILGCAPAIGE 167
Db 61 ywkgtpcarnttbgdcppllemtsyiqgdmvdtgygandftalqfksdvpldicqs 120
Qy 168 HWAKGTACKSRPLSQDCGCPPLKNTVLEGDVMDVDTGYGAMDFSTLQDTKCEVPLDICQS 227
Db 121 lckypdylnmaadpygdmfflrreglfarhfnragvgdkipesylyksgsgrtppg 180
Qy 228 ICKYPDYLQMSADPYGDSMFFCLRRQELFARHFNWNRAGTMGDTVPQSLYIKGTGMASPG 287
Db 181 qnnslenstypfvsgslvtadqfarpfwlkrqaghnng 221
Qy 282 NRASPGSCVYSPSPSGSIVTSDSQLENKPYWLHKAQHNGN 322

RESULT
ID R88268 standard; Peptide; 222 AA.
AC R88268;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 49; papilloma virus; major;
KW capsid protein; plasmid CR148-59; DSM 9134; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
DR N-PSDB; T03500.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 2; 15pp; German.
CC T03500 is 75% homologous to HP-Virus 49, encodes R88268 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid CR148-59 (DSM 9134). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 222 AA;

Query Match 20.3%; Score 754; DB 16; Length 222;
Best Local Similarity 49.6%; Pred. No. 8,626-62;
Matches 111; Conservative 52; Mismatches 50; Indels 11; Gaps 7;
```



Db	1	srghplvgstgthplfnkvktdtenannyivtskdd-rqdsfdpkrvqmfilgcacpige	59
QY	108	GRGQPLGVLSGHPFYKNLDDTESSHAATSNVEDYRDNVSDYKQTQLCILGCAPAIGE	167
Db	60	hwaakpccdadr-gvqkcpplvintviedgdmvdifgninnkntlsanksdvsddivnn	118
QY	168	HWAKGTACKSRPLSQGDCCPPLKNTLVLEGGDMVDYTGAMDFSTLQDTKCEVPLDICQS	227
Db	119	ickypdfikmandlydscffvyrreccvarhffvrggnvgdaipdaavqgdnfnwlpaa	178
QY	228	ICKYPDYQMSADPYGDSMEFFCLLRQELFARHFNNRAGTMGDTVPQS-L-----YIKGTG	281
Db	179	vgaqantlgssiyvptvsgslvstdaqlfnrpfwlrqaghnng	222
QY	282	M-RA-SP-GSCVTSPPSGSIVTSQDLFNKPYLWHLKHAOGHNG	322

RESULT	9
ID	R88269 standard; Peptide; 220 AA.
AC	R88269;
DE	12-JUN-1996 (first entry)
DE	Papilloma virus major capsid protein.
KW	HP-Virus 49; papilloma virus; major;
KW	capsid protein; plasmid VS40-7; DSM 9135; diagnosis;
KW	skin carcinomas; therapy; vaccination.
OS	Papilloma virus.
PN	DE415743-Al.
PD	09-NOV-1994.
PF	04-MAY-1994; 415743.
PR	04-MAY-1994; DE-415743.
PI	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI	de Villiers zur Hausen E, Leigh I, Shumanin V, zur Hausen H;
DR	WPI; 95-383680/50.
DR	N-PSDB; T03501.
PT	DNA encoding peptide(s) of papilloma virus major capsid protein
PT	useful for detecting papilloma virus in skin carcinoma
PS	Claim 7; Fig 3; 15pp; German.
CC	T03501 is 78.5% homologous to HP-Virus 49, encodes R88269 a
CC	peptide of papilloma virus (PV) major capsid protein and is
CC	expressed by the plasmid VS40-7 (DSM 9135). The DNA is useful
CC	in diagnosis, esp. to detect pv in skin carcinomas. Major capsid
CC	proteins encoded by a PV genome contg. the DNA are useful for
CC	therapy and/or vaccination.
SQ	Sequence 220 AA;

Query Match	19.9%	Score 742;	DB 16;	Length 220;
Best Local Similarity	50.0%;	Pred. No. 1.38e-60;		
Matches 111;	Conservative	50;	Mismatches 50;	Indels 11;
				Gaps

Db	1	srqgplvgstgbhplfnkvkxdktdensnyit-mskddrdqtsfdpqqvqmfliigcapcige	59
Qy	108	GRGQPLGVGLSGHPFNKLDLDESSHAATSNVSEDYRDVYKOTQLCILGCAPAICE	167
Db	60	hwdaaapcdadk-gdgkcpplvlnviedgdmvdfginnkttlsanksdvsldivnn	118
Qy	168	HWAKGTACKSRPLSQSGDCCPLELKNVTLEDGMDVDTGYGAMDFTLQDTKCEVPLDICO	227
Db	119	ickypdlkmandiycsfyarrceqvarhffvrgnvgvdrfpnaaavggnnfmlpaa	178
Qy	228	ICKYPDYLQMSADPYGDSMEFFCLRRQLFARHPFNWNRAGMTGDTVPQS-L-----YI--KG	279
Db	179	agaqntlgnsiyvptvsgslvstdaqlnrpfwlrqaghn	220
Qy	280	TGM-RASPGSCVTSPSPGSIYVTSQLFNKPYLWLHKAQGN	320

RESULT	10
ID	R88271 standard; Peptide; 224 AA.
AC	R88271;
DT	12-JUN-1996 (first entry)
DE	Papilloma virus major capsid protein.
KW	HP-Virus 17; papilloma virus; major;

KW	capsid protein; plasmid VS102-4; DSM 9137; diagnosis;
KW	skin carcinomas; therapy; vaccination.
OS	Papilloma virus.
PN	DE4415743-A1.
PD	09-NOV-1995.
PF	04-MAY-1994; 415743.
PR	04-MAY-1994; DE-415743.
PR	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI	de Villiers zur Hausen E, Leigh I, Shananin V, zur Hausen H;
PA	WFI; 95-383680/50.
DR	N-PSDB; T03503.
FT	DNA encoding peptide(s) of papilloma virus major capsid protein
FT	useful for detecting papilloma virus in skin carcinoma
PS	claim 7; Fig 5; 15pp; German.
PS	T03503 is 79% homologous to HP-virus 17, encodes R88271 a
CC	peptide of papilloma virus (PV) major capsid protein and is
CC	expressed by the plasmid VS102-4 (DSM 9137). The DNA is useful
CC	in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC	proteins encoded by a PV genome contg. the DNA are useful for
CC	therapy and/or vaccination.
CC	Sequence 224 AA;
SQ	

```

Query Match      19.68; Score 728; DB 16; Length 224;
Best Local Similarity 49.88; Pred. No. 3.50e-59;
Matches 112; Conservative 49; Mismatches 53; Indels 11; Gaps

Db      1  srgqplvgtsghplnkvrdtens-gnyqavsgddarqntsfdpkqvmfvagcvcpmge 59
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      108  GRQPLGVGLSGHPFNKLLDDESSHAATNSWSEYRDVNVVDYKTKLCILGCAPAIGE 167
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      60  hwdkavcseanqqglcoppelknsviedgmdfignlnnkalsynksadvsldfyn 119
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      168  HWAQGTACKSRPLSQ-GDCPPLEKNTVLGDGDMVDTGYCAMDFSLQTKTKEVPLDICO 226
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120  evckydpdflmadvvgdcaiffarregcyvarhyfvgnvdaipdgavagqdhnylpa 179
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      227  SICKYPDYLOMSADPFGDSMFFCLREQLFAHFWRAGTMDGTVP-----Q-----SLYIKG 279

Db      180  qnaqqqhtlgnslpyptvsgslvtsdaqlnfrpfwlqragqqnng 224
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      280  -TGNRASP-GSCVYSPSGSIYVTSQSLFNKPYTWLHKAGQHNG 322

```

RESULT 11

AC R88274 standard; Peptide: 224 AA.  
AC R88274;  
DE 12-JUN-1996 (first entry)  
DE Papilloma virus major capsid protein.  
KW HPV-Virus 15; papilloma virus; major;  
KW capsid protein; plasmid VS92.1; DSM 9140; diagnosis;  
KW skin carcinomas; therapy: vaccination.  
OS Papilloma virus.  
PN DE415743-A1.  
PD 05-NOV-1995.  
PD 04-MAY-1994; 415743.  
PR 04-MAY-1994; DE-415743.  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;  
DR WPI: 95-383680/50.  
DR N-PSDB; T03506.  
PT DNA encoding peptide(s) of papilloma virus major capsid protein -  
PT useful for detecting papilloma virus in skin carcinoma  
PS Claim 7; Fig 8; 15pp; German.  
CC T03506 is 82.8% homologous to HPV-Virus 15, encodes R88274 a  
CC peptide of papilloma virus (PV) major capsid protein and is  
CC expressed by the plasmid VS92-1 (DSM 9140). The DNA is useful  
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid  
CC proteins encoded by a PV genome contg. the DNA are useful for  
CC therapy and/or vaccination.  
SQ Sequence 224 AA.

Query Match 19.5%; Score 724; DB 16; Length 224;  
Best Local Similarity 49.8%; Pred. No. 8.80e-59;

```
Matches 112; Conservative 48; Mismatches 54; Indels 11; Gaps 5;
Db 1 gqllhvgtaghplfnkvdten-nsyqgdtstddqntsfqpkvqmfvgcapolce 59
Qy 108 GRGQPLGVGLSGHPFYNKLDITSSHAATSNVSEDRDNVSDYKQTQCLIGCAPAIGE 167
Db 60 hwdkapvcdsknagckpplcintviedgmidlgfnginnkvltvskdvsldivn 119
Qy 168 HWAAGTACKSRPLSQ-GCCPPLKNTVLEDDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 226
Db 120 etckypdfltmndvdydcfffarreccyaryhyvrgvgvdalpeavqdknfwlpa 179
Qy 227 SICKPYDYLQMSADPYGDSMFFCLRRQLEFARHFWNRAGTMDTVP-QSL-----LYI 277
Db 180 qbtqgkdlassiyftvsgslvtsdaqlfnrpfwlraggqnnq 224
Qy 278 KGTGMRA-SGSCVSPSGSVITSDSOLFKNPKYWLHKAQGHNNQ 322
RESULT 12
ID R88270 standard; Peptide; 225 AA.
AC R88270;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
KW HP-Virus 25; papilloma virus; major;
KW capsid protein; plasmid VS20-4; DSM 9136; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03502.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 4; 15pp; German.
CC T03502 is 75.6% homologous to HP-Virus 25, encodes R88270 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS20-4 (DSM 9136). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 225 AA;
Query Match 19.5%; Score 724; DB 16; Length 225;
Best Local Similarity 48.9%; Pred. No. 8 80e-59;
Matches 111; Conservative 54; Mismatches 48; Indels 14; Gaps 7;
Db 1 gsglpigsgshplfnkvdtenngtygttkdd-rqnsfapklqmfiigtcpelge 59
Qy 108 GRGQPLGVGLSGHPFYNKLDITSSHAATSNVSEDRDNVSDYKQTQCLIGCAPAIGE 167
Db 60 hwdkapavn-diqgscpplcintvtyggdmadigynlnfkalgqrsvdsldvde 118
Qy 168 HWAAGTACKSRPLSQ-GCCPPLKNTVLEDDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 227
Db 119 ickypdflrmqndvdydcfffarreccyaryhfvrggkpgddlpagaqidagshkneyyl 178
Qy 228 ICKPYDYLQMSADPYGDSMFFCLRRQLEFARHFWNRAGTMDTVP--Q-----LYI 277
Db 179 qasdgsgnslmsnyftlsgslvssdaqlfnrpfwlraggqnnq 225
Qy 278 KGTGMRA-SGSCVSPSGSVITSDSOLFKNPKYWLHKAQGHNNQ 322
RESULT 13
ID R88272 standard; Peptide; 220 AA.
AC R88272;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
```

```
HP-Virus 17; papilloma virus; major;
KW capsid protein; plasmid VS73-1; DSM 9138; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03504.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 6; 15pp; German.
CC T03504 is 73.8% homologous to HP-Virus 17, encodes R88272 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS73-1 (DSM 9138). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 220 AA;
Query Match 19.3%; Score 717; DB 16; Length 220;
Best Local Similarity 49.8%; Pred. No. 4 42e-58;
Matches 110; Conservative 62; Mismatches 62; Indels 7; Gaps 4;
Db 1 gsgpplgvgttghplfnkldrsensaeerlegtssddrr-nisfdpkvqgmfvigtcpelge 59
Qy 108 GRGQPLGVGLSGHPFYNKLDITSSHAATSNVSEDRDNVSDYKQTQCLIGCAPAIGE 167
Db 60 ywdtapvckdagsqlgcpplcintvtyggdmadigfnginnktlfnksdvsdvdivne 119
Qy 168 HWAAGTACKSRPLSQ-GCCPPLKNTVLEDDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 227
Db 120 ickypdfltmndvdydcfffarreccyaryhfvrggavgdllipdvqdhkyyvlpan 179
Qy 228 ICKPYDYLQMSADPYGDSMFFCLRRQLEFARHFWNRAGTMDTVPQSLYIKGTG--MRAS 285
Db 180 ppatlenstyftlsgslvtsdaqlfnrpfwlraggqnnq 220
Qy 286 P-GS---CVTSPSGSVITSDSOLFKNPKYWLHKAQGHNNQ 322
RESULT 14
ID R88275 standard; Peptide; 228 AA.
AC R88275;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
KW HP-Virus 12; papilloma virus; major;
KW capsid protein; plasmid VS75-3; DSM 9141; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03507.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 9; 15pp; German.
CC T03507 is 75.7% homologous to HP-Virus 12, encodes R88275 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS75-3 (DSM 9141). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 228 AA;
Query Match 19.3%; Score 720; DB 16; Length 228;
```

Search completed: Wed Mar 26 08:31:06 1997  
Job time : 108 secs.

Best Local Similarity 50.9%; Pred. No. 2.21e-58;  
Matches 116; Conservative 42; Mismatches 57; Indels 13; Gaps 6;

Db 1 grqqplvgvtsghplfnkvndaenplayraqafstddrqnstsfdkqgmfiigcapcig 60  
QY 108 GKGQPLGVLSGHPFYNKLDDESHATSNV-SEDRVNDNSVDYKQQLCILGCPAIG 166  
Db 61 ehwdgercagannengcoppiklvnsiqdgmadiygnlnfrtlgenrdsvaldivn 120  
QY 167 EHWAGTACKSRPLSGDCPPLELANTVLEGDMDVDTGCGAMDFTLQDTKCEVPLDICQ 226  
Db 121 etckypdfikmndlygdscofffarrecyarrhffvrgkagddlpagaqdagtykndfy 180  
QY 227 SICKYPDYQLMSADPYGSMFFCLRREQLFARHFNWRAAGTMDGTVP--Q-----LY 276  
Db 181 ipgasgqkngnsmypftvsgslvssdaqlfnrpfwlgragqgnng 228  
QY 277 IKGT-G-MRASPSCVYSPSPSGSIVTSDSQLFNRPYWLHRAQGHNG 322

## RESULT 15

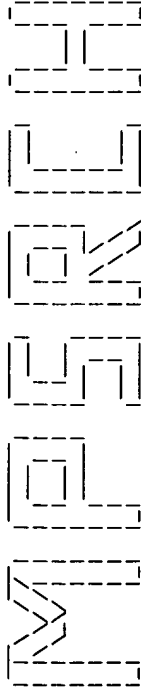
ID R32057 standard; Protein; 88 AA.  
AC R32057;  
DE 07-JUN-1993 (first entry)  
KW Rabbit serum reactive HPV16 L1 epitope 250-337.  
KW B; epitope; human; papilloma; virus; HPV-16; L1; antisera; ELISA; A1;  
KW antigenic index; flexibility; accessibility; hydrophilicity; mice;  
KW immunise; HPV16 capsid; papilloma virus like particle; VLP; diagnosis;  
KW vaccine.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 18..20  
FT /note= "Epitope 267-269"  
PN W09302184-A.  
PD 04-FEB-1993.  
PF 20-JUL-1992; AU0364.  
PR 19-JUL-1991; AU-007322.  
PA (CLSC-) CLS LTD.  
PA (UYQU) UNIV QUEENSLAND.  
PI Frazer I, Zhou J;  
DI WPI: 93-058783/07.  
PT Prodn. of papilloma virus-like particles - which contain L1 and  
PT L2 proteins, useful as vaccine against papilloma virus infections  
PT and in diagnosis  
PS Claim 39; Table 1; 63pp; English.  
CC The sequences given in R32054-73 are B epitopes which are derived  
CC from the human papilloma virus-16 (HPV-16) L1 protein. These  
CC epitopes were isolated by screening antisera from immunised animals  
CC against a set of overlapping HPV-16 L1 peptides by ELISA. Possible  
CC antigenic regions were identified using an antigenic index (AI) on the  
CC basis of chain flexibility, high accessibility and a high degree of  
CC hydrophilicity. A region with an AI value of over 1.5 was regarded as  
CC a predicted B epitope. Five regions identified in this manner were  
CC found to be within the 22 peptides to which major reactivity was  
CC seen with antisera from mice immunised with synthetic HPV16 capsids.  
CC These epitopes can be used to in the production of papilloma virus  
CC like particles (VLPs) for use as diagnostic agents and as components  
CC of vaccines for use with papilloma virus infections.  
SQ Sequence 88 AA;

Query Match 12.5%; Score 466; DB 6; Length 88;  
Best Local Similarity 65.9%; Pred. No. 3.33e-33;  
Matches 58; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

Db 1 ictsickypdyikmvsepygdsilffylreqmfvrhlfnragtgvnvpddlylkgsgst 60  
QY 224 ICQSIKCKYDYQLMSADPYGSMFFCLRREQLFARHFNWRAAGTMDGTVPQSLYIKGTGM 283  
Db 61 anlansnyftpsgsmvtsdaqlfnkpy 88  
QY 284 ASPGSCVYSPSPSGSIVTSDSQLFNRPY 311



\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Mar 27 14:05:45 1997; MasPar time 751.98 Seconds

Tabular output not generated. 893.808 Million cell updates/sec

Title: >US-08-409-122-1  
Description: (1-1524) from US08409122.seq

Perfect Score: 1524

N.A. Sequence:

Comp: 1 ATGGCTTTGTGGGGCCCTAG.....GTGACGTGCGGAGTAA 1524

TACCGAACACCCCGGATC.....CACATGACACGCTCTTCATT

Scoring table:

TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 630489 seqs, 220513910 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1-EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9-EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134 135:EST135  
136:EST136 137:EST137 138:EST138 139:EST139 140:EST140  
141:EST141 142:EST142 143:EST143 144:EST144 145:EST145  
146:EST146 147:EST147 148:EST148 149:EST149 150:EST150  
151:EST151 152:EST152 153:EST153 154:EST154 155:EST155  
156:EST156 157:EST157 158:EST158 159:EST159 160:EST160  
161:EST161 162:EST162 163:EST163 164:EST164 165:EST165  
166:EST166 167:EST167 168:EST168 169:EST169 170:EST170  
171:EST171 172:EST172 173:EST173 174:EST174 175:EST175  
176:EST176 177:EST177 178:EST178 179:EST179 180:EST180  
181:EST181 182:EST182 183:EST183 184:EST184 185:EST185  
186:EST186 187:EST187 188:EST188 189:EST189 190:EST190  
191:EST191 192:EST192 193:EST193 194:EST194 195:EST195  
196:EST196 197:EST197 198:EST198 199:EST199

Database:

173:enSTS1 174:enSTS2

Statistics:

Mean 11.323; Variance 1.974; scale 5.736

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	25	1.6	391	74	Yf36h09.r1 Homo sapie	9.38e-08
C 2	24	1.6	418	160	Yf76e04.r1 Homo sapie	2.21e-06
C 3	24	1.6	418	141	Yf76e04.r1 Homo sapie	2.21e-06
C 4	24	1.6	418	141	Yf76e04.r1 Homo sapie	2.21e-06
C 5	25	1.6	442	28	Yf01c07.r1 Homo sapie	9.38e-08
C 6	23	1.5	477	9	Ym07d11.s1 Homo sapie	4.66e-05
C 7	22	1.4	162	107	T19069	8.78e-04
C 8	21	1.4	205	2	BNAF1787	1.46e-02
C 9	21	1.4	231	93	R76852	1.46e-02
C 10	21	1.4	243	153	W20635	1.46e-02
C 11	21	1.4	243	170	MM63512	1.46e-02
C 12	21	1.4	248	128	T99295	1.46e-02
C 13	22	1.4	250	79	R26164	8.78e-04
C 14	22	1.4	257	151	W17718	8.78e-04
C 15	22	1.4	257	171	MM7189	8.78e-04
C 16	21	1.4	260	30	H80328	1.46e-02
C 17	21	1.4	263	44	HSCZOD022	1.46e-02
C 18	21	1.4	267	79	R25478	1.46e-02
C 19	22	1.4	271	48	HDM292B04B	8.78e-04
C 20	21	1.4	274	44	HSCZOD012	1.46e-02
C 21	21	1.4	301	95	R81312	1.46e-02
C 22	22	1.4	332	47	HOM112B01A	1.46e-02
C 23	21	1.4	356	22	H53416	8.78e-04
C 24	21	1.4	357	46	HDM014B03A	8.78e-04
C 25	22	1.4	360	94	R78356	8.78e-04
C 26	21	1.4	363	136	HS4053YH9	1.46e-02
C 27	21	1.4	382	111	T37006	1.46e-02
C 28	21	1.4	390	148	W07840	1.46e-02
C 29	21	1.4	400	104	T04435	1.46e-02
C 30	22	1.4	400	133	G11249	8.78e-04
C 31	22	1.4	406	26	H64976	8.78e-04
C 32	22	1.4	407	79	R28023	8.78e-04
C 33	21	1.4	411	150	HS313320	1.46e-02
C 34	22	1.4	415	11	H16705	8.78e-04
C 35	21	1.4	418	82	R39694	1.46e-02
C 36	22	1.4	422	83	R40554	8.78e-04
C 37	22	1.4	423	146	W01843	8.78e-04
C 38	22	1.4	423	166	HS843329	8.78e-04
C 39	21	1.4	430	133	T43130	1.46e-02
C 40	21	1.4	431	91	R69853	1.46e-02
C 41	21	1.4	441	96	R83692	1.46e-02
C 42	21	1.4	455	157	AT26618	1.46e-02
C 43	22	1.4	473	52	N20104	8.78e-04
C 44	21	1.4	551	146	W02574	1.46e-02
C 45	21	1.4	579	131	G07099	1.46e-02

ALIGNMENTS

1 RI0310 391 bp mRNA EST 06-APR-1995  
LOCUS Yf36h09.r1 Homo sapiens cDNA clone 128993 5'  
DEFINITION Yf36h09.r1 Homo sapiens cDNA clone 128993 5'  
ACCESSION Yf36h09.r1 Homo sapiens cDNA clone 128993 5'  
NID G762266  
KEYWORDS EST  
SOURCE human clone-128993 library-Soares fetal liver spleen INFLS  
vector-pir73D (Pharmacia) with a modified polylinker host-DH10B  
(ampicillin resistant) primer-M13R1 Rsite1-Pac 1 Rsite2-Eco RI  
Liver and spleen from a 20 week-post conception male fetus. 1st  
strand cDNA was primed with a Pac I - Oligo(dT) primer [5']





Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

# TITLE JOURNAL COMMENT

GDB: G00-419-911  
Contact: Wilson, R.K.  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 347

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Location/Qualifiers

## FEATURES

source

1..477

/organism="Homo sapiens"

/clone="47370"

/note="human"

BASE COUNT 132 a 79 c 92 g 170 t 4 others

## ORIGIN

Query Match 1.5%; Score 23; DB 9; Length 477;

Best Local Similarity 64.9%; Pred. No. 4.66e-05;

Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Db 144 taaagtgtttcttcacactcagcttttttaataagaacattgaaatcaacgaagtgt 203

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 1200 TAAATACTGCTATTCACTACTGATGATATAGGACATACATCTGCAGTTAAGTAAGT 1141

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 204 agttaactaaaaataaa 220

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 1140 ACATAACTGAAAAATAA 1124

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7 T19069 162 bp mRNA EST 08-AUG-1995

LOCUS f09005s Homo sapiens cDNA clone f09005 5' end.

DEFINITION T19069

ACCESSION T19069

NID G60112

KEYWORDS EST.

SOURCE human clone-f09005 library-Testis 2 vector-PGEM 5zf(+)-primer-SP6

Rsitel-Sall Rsitel2-NotI mRNA was prepared from human testis of a 27

years old man. cDNA was prepared using a 15mer oligo dr anchored by

two degenerated bases at its 3' end and containing a NotI site at

1cs 5' end. The cDNA was cloned between Sall and NotI sites of PGEM

5zf(+). The 3' end is at the NotI site. The Sall site is lost

during the cloning procedure. cDNA corresponding to abundant

species were eliminated from this library.

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;

Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 162)

Pawlak, A., Touseaint, C., Levy, I., Bulle, F., Poyard, M., Barouki, R.

and Guellaen, G.

Characterization of a large population of mRNAs from human testis

Genomics 26, 151-158 (1995)

Contact: Guellaen G

Unite INSERM 99

INSERM

Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France

Tel: (33)149813530

Fax: (33)14980908

Email: guellaen@cit12.fr.

Location/Qualifiers

1..162

FEATURES

source

/organism="Homo sapiens"

/clone="f09005"

/note="human"

BASE COUNT 24 a 66 c 45 g 25 t 2 others

## ORIGIN

Query Match 1.4%; Score 22; DB 107; Length 162;

Best Local Similarity 72.0%; Pred. No. 8.78e-04;

Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 50 gccgcgcgtcagggctgctccagctgcctcctcctcctcctacactgcc 99

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1445 GCCCTCGAAGACCTTCTCTCATCTGCCACTACGCTTCTTAACCTGCC 1494

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8 BNAF1787 205 bp mRNA EST 13-OCT-1995

LOCUS Brassica campestris (clone F1787) expressed sequence tag (EST)

DEFINITION L47916

ACCESSION L47916

NID g1019530

KEYWORDS EST; expressed sequence tag.

SOURCE Brassica campestris (clone: F1787) flower bud cDNA to mRNA.

ORGANISM Brassica campestris

Eukaryotae; Mitochondrial eukaryotes; Viridiplantae;

Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;

Magnoliopsida; Capparales; Brassicaceae; Brassica.

1 (bases 1 to 205)

Liu, C.O., Kim, M.G. and Cho, M.J.

Generation of expressed sequence tags from random sequencing of

Chinese cabbage (Brassica campestris L. ssp. pekinensis) flower bud

cDNA clones

Unpublished (1995)

Location/Qualifiers

1..205

/organism="Brassica campestris"

/clone="F1787"

/dev\_stage="flower bud"

/sequenced\_mol="cDNA to mRNA"

1..205

/standard\_name="expressed sequence tag"

/note="putative"

BASE COUNT 45 a 25 c 85 g 50 t

ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 205;

Best Local Similarity 74.4%; Pred. No. 1.46e-02;

Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 43 tggagaaggaattggctatggctgagtggttcaggtgttgccatt 85

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 309 TGGAGTGGAAATTGGCCGTTGTCAGCCCTTATAGGTTTGGCCTT 351

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9 R76852 231 bp mRNA EST 06-JUN-1995

LOCUS Y163d10.s1 Homo sapiens cDNA clone 143923 3'.

DEFINITION R76852

ACCESSION R76852

NID g851484

KEYWORDS EST.

SOURCE human clone-143923 library-Soares placenta Nb2HP vector-pt7r3D

(Pharmacia) with a modified polylinker host-DH10B (ampicillin

resistant) primer-Promega -21ml3 Rsitel-Not I Rsitel2-Eco RI Female

placenta obtained at birth (full term). 1st strand cDNA was primed

with a Not I - oligo(dt) primer [5'

AACCTGAAGATTCGCGCCGACAGGATTTTTTTTTTTT 3'], double-stranded

cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Not I and Eco RI sites of the modified pT73

vector. Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bernaldo.

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

1..162

FEATURES

source





```

FT      /clone="348711"
FT      /clone_lib="Soares mouse p3NMF19.5"
FT      /dev_stage="19 weeks"
FT      /lab_host="DH10B (ampicillin resistant)"
FT      <1..>243
SQ      Sequence 243 BP; 95 A; 51 C; 37 G; 60 T; 0 other;

Query Match
Best Local Similarity 88.98; Pred No. 1.46e-02; Length 243;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db      125 aactccaagaacacttgcctttaa 151
      |||||
Cp      1379 AGGTCAGAGAACTTTCCTTTAA 1353

RESULT 12
LOCUS   T99295          248 bp      mRNA      EST      31-MAR-1995
DEFINITION yes3908.r1 Homo sapiens cDNA clone 122462 5'.
ACCESSION T99295
NID       9749032
KEYWORDS EST.
SOURCE    human clone-122462 library-Soares fetal liver spleen INFLS
vector-pt7T3D (Pharmacia) with a modified polylinker host-DH10B
(ampicillin resistant) primer-M13R1 Rsite1-Pac I Rsite2-Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo(dT) primer [5',
AATGGGAAGAATTAATTAAGATTTTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT7T3
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaudo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 248)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

FEATURES
source
1..248
/organism="Homo sapiens"
/clone="122462"
/notes="human"

BASE COUNT 94 a 62 c 39 g 53 t

ORIGIN
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 237
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
source
1..248
/organism="Homo sapiens"
/clone="122462"
/notes="human"

BASE COUNT 94 a 62 c 39 g 53 t

ORIGIN
Query Match 1.48; Score 21; DB 128; Length 248;
Best Local Similarity 92.08; Pred. No. 1.46e-02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db      8 aatcaactgaactcaattcttca 32
      |||||
Cp      1128 AATAAAGTCAATCATATCTTCA 1104

RESULT 13

```

```

LOCUS   R26164          250 bp      mRNA      EST      24-APR-1995
DEFINITION Yh39e12.s1 Homo sapiens cDNA clone 132142 3'.
ACCESSION R26164
NID       9782299
KEYWORDS EST.
SOURCE    human clone-132142 library-Soares placenta Nb2Hp vector-pt7T3D
(Pharmacia) with a modified polylinker host-DH10B (ampicillin
resistant) primer-promega-2nl3 Rsite1-Not I Rsite2-Eco RI Female
placenta obtained at birth (full term) . 1st strand cDNA was primed
with Not I - oligo(dT) primer [5',
AATGGGAAGAATTCGCGCCGCGAATTTTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT7T3
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaudo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

FEATURES
source
1..250
/organism="Homo sapiens"
/clone="132142"
/notes="human"

BASE COUNT 100 a 36 c 25 g 89 t

ORIGIN
Query Match 1.48; Score 22; DB 79; Length 250;
Best Local Similarity 75.0%; Pred. No. 8.78e-04;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db      74 ataacatttcattcacgtataacacatgaataattataa 117
      |||||
Cp      1166 ATAACATCGCATTAAGTAATAGTACATAAATGAAATATAA 1123

RESULT 14
LOCUS   W17718          257 bp      mRNA      EST      29-APR-1996
DEFINITION mb577f05.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 335457
ACCESSION 5'.
NID       W17718
KEYWORDS g1292119
SOURCE    EST.
ORGANISM house mouse.
Mus musculus
Eucaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 257)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

```





\*\*\*\*\*  
W P S R L H  
\*\*\*\*\*  
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:26:09 1997; MasPar time 14.21 Seconds  
Tabular output not generated. 661.355 Million cell updates/sec

Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pap  
Perfect Score: 3721  
Sequence: 1 MALWRPSDNTVLPSPVAR.....APSATTSSKPAKRVVRARK 507

Scoring table: PAM 150  
Gap 11

Searched: 52205 seqs, 18531385 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot33  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10

Statistics: Mean 49.368; Variance 79.777; scale 0.619

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3684	99.0	568	9	VLL_HPV18	PROBABLE L1 PROTEIN. 0.00e+00
2	3398	91.3	539	9	VLL_HPV15	PROBABLE L1 PROTEIN. 0.00e+00
3	3052	82.0	505	9	VLL_HPVME	PROBABLE L1 PROTEIN. 0.00e+00
4	3030	81.4	505	9	VLL_HPV39	PROBABLE L1 PROTEIN. 0.00e+00
5	2751	73.9	531	9	VLL_HPV10	PROBABLE L1 PROTEIN. 0.00e+00
6	2723	73.2	532	9	VLL_HPV03	PROBABLE L1 PROTEIN. 0.00e+00
7	2690	72.3	503	9	VLL_HPV26	PROBABLE L1 PROTEIN. 0.00e+00
8	2684	72.1	508	9	VLL_HPV30	PROBABLE L1 PROTEIN. 0.00e+00
9	2656	71.4	499	9	VLL_HPV33	PROBABLE L1 PROTEIN. 0.00e+00
10	2644	71.1	534	9	VLL_HPV36	PROBABLE L1 PROTEIN. 0.00e+00
11	2641	71.0	501	9	VLL_HPV1	PROBABLE L1 PROTEIN. 0.00e+00
12	2643	71.0	504	9	VLL_HPV51	PROBABLE L1 PROTEIN. 0.00e+00
13	2624	70.5	504	9	VLL_HPV31	PROBABLE L1 PROTEIN. 0.00e+00
14	2600	69.9	502	9	VLL_HPV35	PROBABLE L1 PROTEIN. 0.00e+00
15	2592	69.7	510	9	VLL_HPV2A	PROBABLE L1 PROTEIN. 0.00e+00
16	2589	69.6	531	9	VLL_HPV16	PROBABLE L1 PROTEIN. 0.00e+00
17	2567	69.0	510	9	VLL_HPV57	PROBABLE L1 PROTEIN. 0.00e+00
18	2567	69.0	594	9	VLL_HPV27	PROBABLE L1 PROTEIN. 0.00e+00
19	2549	68.5	524	9	VLL_HPV34	PROBABLE L1 PROTEIN. 0.00e+00
20	2550	68.5	528	9	VLL_HPV38	PROBABLE L1 PROTEIN. 0.00e+00
21	2546	68.4	500	9	VLL_HPV6B	PROBABLE L1 PROTEIN. 0.00e+00
22	2547	68.4	505	9	VLL_HPV07	PROBABLE L1 PROTEIN. 0.00e+00

23	2540	68.3	499	9	VLL_HPV33	PROBABLE L1 PROTEIN. 0.00e+00
24	2542	68.3	502	9	VLL_PCPIV1	PROBABLE L1 PROTEIN. 0.00e+00
25	2542	68.3	505	9	VLL_HPV40	PROBABLE L1 PROTEIN. 0.00e+00
26	2526	67.9	529	9	VLL_HPV52	PROBABLE L1 PROTEIN. 0.00e+00
27	2520	67.7	499	9	VLL_HPV13	PROBABLE L1 PROTEIN. 0.00e+00
28	2512	67.5	501	9	VLL_HPV11	PROBABLE L1 PROTEIN. 0.00e+00
29	2484	66.8	503	9	VLL_HPV32	PROBABLE L1 PROTEIN. 0.00e+00
30	2475	66.5	502	9	VLL_HPV42	PROBABLE L1 PROTEIN. 0.00e+00
31	1992	53.5	508	9	VLL_HPV1A	PROBABLE L1 PROTEIN. 0.00e+00
32	1981	53.2	546	9	VLL_HPV19	PROBABLE L1 PROTEIN. 0.00e+00
33	1972	53.0	516	9	VLL_HPV65	PROBABLE L1 PROTEIN. 0.00e+00
34	1966	52.8	517	9	VLL_HPV25	PROBABLE L1 PROTEIN. 0.00e+00
35	1960	52.7	509	9	VLL_HPV49	PROBABLE L1 PROTEIN. 0.00e+00
36	1962	52.7	518	9	VLL_HPV14	PROBABLE L1 PROTEIN. 0.00e+00
37	1950	52.4	505	9	VLL_CRPVK	PROBABLE L1 PROTEIN. 0.00e+00
38	1950	52.4	507	9	VLL_HPV63	PROBABLE L1 PROTEIN. 0.00e+00
39	1949	52.4	507	9	VLL_HPV17	PROBABLE L1 PROTEIN. 0.00e+00
40	1951	52.4	516	9	VLL_HPV04	PROBABLE L1 PROTEIN. 0.00e+00
41	1945	52.3	514	9	VLL_HPV08	PROBABLE L1 PROTEIN. 0.00e+00
42	1938	52.1	517	9	VLL_HPV12	PROBABLE L1 PROTEIN. 0.00e+00
43	1932	51.9	507	9	VLL_HPV15	PROBABLE L1 PROTEIN. 0.00e+00
44	1923	51.7	507	9	VLL_HPV09	PROBABLE L1 PROTEIN. 0.00e+00
45	1917	51.5	514	9	VLL_HPV47	PROBABLE L1 PROTEIN. 0.00e+00

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	568 AA.
ID VLL_HPV18			
AC P06794;			
DT 01-JAN-1988 (REL. 06, CREATED)			
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)			
DE 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)			
DE PROBABLE L1 PROTEIN.			
GN L1.			
OS HUMAN PAPILLOMAVIRUS TYPE 18.			
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 87283882.			
RA COLE S.T., DANOS O.;			
RL J. MOL. BIOL. 193:599-608(1987).			
RN [2]			
RP SEQUENCE OF 363-406 FROM N.A.			
RX MEDLINE; 92407963.			
RA CHAN S.Y., BERNARD H.U., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;			
RL J. VIROL. 66:5714-5725(1992).			
DR EMBL; X05015; G60983;			
DR EMBL; M96287; G333172;			
DR EMBL; A06329; G413678;			
DR PIR; A26251; P1WL18.			
KW LATE PROTEIN.			
FT CONFLICT 384 384 V -> I (IN REF. 2).			
SQ SEQUENCE 568 AA; 63623 MW; 8D0D12FC CRC32;			
Query Match 99.0%; Score 3684; DB 9; Length 568;			
Best Local Similarity 99.0%; Pred.No. 0.00e+00;			
Matches 502; Conservative 1; Mismatches 4; Indels 0; Gaps 0;			
Db 62 malwrpsdntvlpppsvarvntddvvtptsfyhgagssrlltvgnpfyrpagggnkq 121			
QY 1 MALWRPSDNTVLPSPVARVNTDDVVTSTSYFHAGSSRLTVCNPFYRVPAGGNKQ 60			
Db 122 dipkvsaygyrvfrvlpdpnkfgldpdtstynpqrllwacagvagr9qplgvlsgn 181			
QY 61 DIPKVSAYGYRVFRVLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLGVLSGH 120			
Db 182 pfynklddteesshaatsnsvsedvrdnsvdykdtqcllgcapaicehwakgtacksrpl 241			
QY 121 PFYNKLDDESSHAATSNVSEDRDNDVSDYKDTQCLLGCAIPAIGHWAKGTACKSRPL 180			
Db 242 sqgdcpplelknvtvledgdmvdtggamdfstlqdkcvepldicsickypdylgmsad 301			

[illegible]

QY	358	ATKFKQYSRHVEEYDLQFQLCTITITADVMYSIHSNNSILEDWNFGVPPPTTSLVD	417
Db	447	tyrfvqvavtcqkdttppekqdydklxfwtvdiikefssdlidqyplgrkflvqaglr	506
QY	418	TYRFVQSVAITCOKDAAPENKDPYDKLAFVNDLKERFSLDLDOYPLGRKFLVQAGLR	477
Db	507	rptgprkrpaastastastaspakrvirskk	539
QY	478	KPTIGPKRSAPSA-T--TSSKPAKRVVRANK	507

RESULT	3	
ID	VLL_HPVME	STANDARD; PRT; 505 AA.
AC	P27564;	
DT	01-AUG-1992 (REL. 23, CREATED)	
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)	
DT	01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)	
DE	PROBABLE L1 PROTEIN.	
GN	L1	
OS	HUMAN PAPILLOMAVIRUS TYPE ME180	
OC	VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.	
RN	[1]	
RA	SEQUENCE FROM N.A.	
RA	MEDLINE; 91374616.	
RL	REUTER S., DELLOS H., KAHN T., HOFMANN B., ZUR HAUSEN H., SCHWARZ E.;	
RL	J. VIROL. 68:5364-5368(1991).	
DR	EMBL; M73258; ; NOT_ANNOTATED_CDS.	
DR	PIR; B40509; PIWLPR.	
DR	LATE PROTEIN.	
SK	SEQUENCE 505 AA; 56805 MW; 32FD93FE CRC32;	

Query Match	82.0%;	Score 3052;	DB 9;	Length 505;
Best Local Similarity	76.7%;	Pred. No. 0.00e+00;		
Matches	386;	Conservative	76;	Mismatches 41; Indels 0; Gaps 0;

Db	1	malwrssdmnviappsvakvntddyttrtglyyagtslltvgdhpyfkypmsgggkq	60
QY	1	MALWRSPDNTVYLPSPVARVNTDDYVTRTSIFVHAGSSRLLTGVNPYFRVPAGGNKQ	60
Db	61	dipkvsayqyrvfrislpdnpkfslpestlynpdtkrlvwacvqveigrqplvgylsg	120
QY	61	DIPKVSAYQYRVFRVGLPDPNKFGLPDNSYINPOTRLVWACAGVEIGRQPLVGLSGH	120
Db	121	plynrladtenspfnsknpxkdsrnmvsdykgtqlcligcvepaigehwakgskcspnv	180
QY	121	PFYKLLDDTSSHAATSNVSEDRDNRVSDYVKTQLCLIGCAPAIGEHWAKGTACKSRPL	180
Db	181	qpgdcpcpleivnptqgdmdtdtgamdfstlqetkeevpldicsvckypdylqmead	240
QY	181	QPGDCPPLELKNVILDEGDMDVDTGGMDFSTLQTKCEVPLDICSICKYDPYLOXMSD	240
Db	241	vygdamfclrrqlfarhfnrgmgvgdtkptelylkgtdirdspesvyvapspsgm	300
QY	241	PYGDSMFCURRQLFARHFNRAKMGTDIVQSYIYKGTGRASPCSCYVSPSGSIV	300
Db	301	sdselgfnkpywlhkeqhgngicwhnqlflvtvdttrstnftlsttstesapnlydnk	360
QY	301	TSDSGLFNKPYWLHKAQHGNGICWHNQLFVTVDTTRSTNLTICASTQSPVPGGYDATK	360
Db	361	fkeylrhveeydlqiflqcltistdmsyhtnmpalldwnfgvappesaslvdtyr	420
QY	361	FKQYSRRHVEEYDLQFQLCTITITADVMYSIHSNNSILEDWNFGVPPPTTSLVDYR	420
Db	421	yqsaatcckdpapthkdydglfnfwnnikexfsseldqfplgrkflvqagvrrpt	480
QY	421	FVQSVAITCOKDAAPENKDPYDKLAFVNDLKERFSLDLDOYPLGRKFLVQAGLRKPT	480
Db	481	igprkrpatatstaskhkriv	503
QY	481	IGPKRSAPSAATSSKPAKRVV	503

RESULT	4	
--------	---	--

```

ID VLI_HPV39 STANDARD; PRT; 505 AA.
AC P24838;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 39.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91135017.
RA VOLPERS C., STRECK R.E.;
RL VIROLOGY 181:419-423(1991).
DR EMBL; M62849; G463192; -.
KW LATE PROTEIN.
SQ SEQUENCE 505 AA; 56604 MW; F07CFCE CRC32;

Query Match 81.4%; Score 3030; DB 9; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 389; Conservative 76; Mismatches 40; Indels 2; Gaps 2;

Db 1 namrsdsmylpppsvkvntdyvtrtgyyaagsslltvghpyfkgmnggrtk 60
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNPYFRVPAGGNKQ 60
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 dipkvsaygyvrvfrlvpdknfsldpnslynpetgrlrvwacvgevgrgplvglsgh 120
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 plynrqddtens-pfssttnkdsdnvadvkqtlcligcvpaigehwkgkackpnnv 179
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 PFYNKLDDESSHAATSNVSDVRNVDYKQTLCLIGCAPAIGEHWAGTACKSRPL 180
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 stgdcpllelvntpiedgmdtdtygandfgalgetksevpldicqscickypdylqmsad 239
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 SQGDCPPELEKNTVLEDDGMDVDTGYGDMDFSTLQDTRCEVPLDICQSCICKYPDYLQMSAD 240
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 vvgdsmfclredlfarfhnrgmvgdipadqlykgtidiranpgssvycpspsgm 299
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 PYGDSMFCLAREQFAHFHWRNAGTMDTVPOSLYIKGTGMRASPGSCVYSPSPSGSIV 300
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 tsdqlfnkpywlhkaqngnngicwhnglflitvvdtrstntflstlessipstypsk 359
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 TSDQLFNKPYWLHKAQNGNNGICWHNGLFVTVVDTRSTNTLTICASQSPVPGQYDATK 360
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 fkeytrhveeydlqfqlctvtltdvmsyihmssildnwnfapvappssalvdyr 419
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 FKQYSRRHVEEYDLQFIFQLCTITITADVMSYIHSNMSSILEDNWFVPPPTTSLVDYR 420
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 ylgaaaitcdapapektdpydglkfnvdrlkfsldqfplgrkflilqarvrrpt 479
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 FVQSVAITCQDAAPAEKNKDPYDLKFNWVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 480 igprkrpaastsses-akhrkrvsk 505
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 481 IGPRKRSAPSATTSSKPAKRVVRAR 507
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
ID VLI_HPV10 STANDARD; PRT; 531 AA.
AC P36732;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 10.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94265501.

Query Match 73.2%; Score 2723; DB 9; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

```

```

RA DELIUS H., HOFMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-29 IS THE INITIATOR.
DR EMBL; X74455; G396908; -.
DR PIR; S36537; S36537.
KW LATE PROTEIN.
SQ SEQUENCE 531 AA; 59011 MW; 5A30125F CRC32;

Query Match 73.9%; Score 2751; DB 9; Length 531;
Best Local Similarity 67.9%; Pred. No. 0.00e+00;
Matches 342; Conservative 103; Mismatches 57; Indels 2; Gaps 2;

Db 29 malwrsdnlylppptvskvstddvtrtnlyyyagtslltvghpyfipkssnnkv 88
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNPYFRVPAGGNKQ 60
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 89 dvpkvsafyrvfrvrlpdpnkfplpdaerlynpdaerlywactvevgrgplvglsgh 148
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 149 plynkledtenslahpignqsdnsldnkdqtlcligctppmgehkwgktpcrpp- 207
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 PFYNKLDDESSHAATSNVSDVRNVDYKQTLCLIGCAPAIGEHWAGTACKSRPL 180
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 208 agqdcplelitspigdmdvtyggamdfalqlnksdvpldicqstckypdylgmaae 267
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 SQGDCPPELEKNTVLEDDGMDVDTGYGDMDFSTLQDTRCEVPLDICQSCICKYPDYLQMSAD 240
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 268 pygdsmfylrredlfarfhnrasavgdalpdtdfllksnggrdvgsavyspbgsmv 327
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 PYGDSMFCLAREQFAHFHWRNAGTMDTVPOSLYIKGTGMRASPGSCVYSPSPSGSIV 300
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 328 tseqlfnkpywlhkaqngnngicwhnglflitvvdtrstntmclvpsseapattydatk 387
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 TSDQLFNKPYWLHKAQNGNNGICWHNGLFVTVVDTRSTNTLTICASQSPVPGQYDATK 360
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 388 fkeytrhveeydlqfqlctvtltdvmsyihmssildnwnfgitlpstledtyr 447
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 FKQYSRRHVEEYDLQFIFQLCTITITADVMSYIHSNMSSILEDNWFVPPPTTSLVDYR 420
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 448 flssaatcqdtpptekqbpaklnfwdvdkldrfldslsqfplgrkflilqlgvrrsa 507
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 FVQSVAITCQDAAPAEKNKDPYDLKFNWVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 508 vsrvkrpatsatgt-aakrkrk 530
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 481 IGPRKRSAPSATTSSKPAKRVVR 504
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 6
ID VLI_HPV03 STANDARD; PRT; 532 AA.
AC P36731;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 3.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94265501.
RA DELIUS H., HOFMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-29 IS THE INITIATOR.
DR EMBL; X74462; G397012; -.
DR PIR; S36554; S36554.
KW LATE PROTEIN.
SQ SEQUENCE 532 AA; 59194 MW; EFADFCL3 CRC32;

Query Match 73.2%; Score 2723; DB 9; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

```

Db 29 malwrsdnlvlpptvskvlatdyvtrtniyyagssrllltvghpfaipksnsmk 88  
QY 1 MALWRSDNTVLPPTVSKVARVNTDDYVTRTSIFVHAGSSRLLTGVPYFRVPAGGNKQ 60  
Db 89 dipkvaafyrvfvrldpdknfgldarfyndpdaerlvwactvgvgrgplvglgsh 148  
QY 61 DIPKVASAYQYRVFVRLDPDKNFGLDARFYNDPDAERLVWACTVGVGRGPLVGLSGH 120  
Db 149 plnyklldtensatnahdickqrdnsvdnyvdykqtolcilcapaigehwakgtacksrpl 207  
QY 121 PFYNKLDDESSHAATNSVSEDVRDNYVDYKQTOLCILCAPAIGEHWAKGTACKSRPL 180  
Db 208 spdcplleltatpdcgndvctgandfnglnsknsdvpldictckypdylgmaae 267  
QY 181 SQGDCPPELNTVLEQDMVDYIGYAMDFTSLQDTKEVPLDICISICKYDPYLQMSAD 240  
Db 268 pygdemffylrkeqlfarhfnraghngicwgnqlfvtcvdtrstnlticstasleat 327  
QY 241 PYGDSMFFCLRRQELFAHFHFNWNRAGTMDYVQSLYIKG--TCMRASPGSCVYSPSPSG 297  
Db 328 smvtsetqlfknkpylrragghngicwgnqlfvtcvdtrstnlticstasleat--yd 385  
QY 298 SYVTSQSLFNKPYLRRAGGHNGICWGNQLFVTVVDTTSTRNLICASTQSPVPGQYD 357  
Db 386 atkfeylrhgeeyldqfqlckvltlpeimaylhtcmstlledwnfgltltpstaled 445  
QY 358 ATKFEQYSHVEYDLQFIFQLCTITLTADVMXSIHNSNILEDNFVGPPTTSLVD 417  
Db 446 tyrltssaitckdaptckqpyaklnfwdvdkrdfrldisqfplgrkflmglygt 505  
QY 418 TYRFVQSVAITCQKDAAPAKENKDPYDKLFKNVNDLKEFSLDLQYPLGRKFLVQAGLRR 477  
Db 506 rssiisvkrksa--ttsrtaaaakrtk 531  
QY 478 KPIGPKRSAPSATTSSKPAKRVRR 504  
RESULT 7  
ID VLI\_HPV26 STANDARD; PRT; 503 AA.  
AC P36735;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE PROBABLE L1 PROTEIN.  
GN L1.  
OS HUMAN PAPILLOMAVIRUS TYPE 26.  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
RN [1]  
RX MEDLINE; 94265501.  
RA DELIUS H., HOFMANN B.;  
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
RN [2]  
RX MEDLINE; 94403963.  
RA CHAN S.Y., BERNARD H.D., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;  
RL J. VIROL. 66:5714-5725(1992).  
DR EMBL; X74474; G396963; -.  
DR PIR; S36549; S36549.  
KW LATE PROTEIN.  
SQ SEQUENCE 503 AA; 56328 MW; 5E31786A CRC32;  
Query Match 72.3%; Score 2690; DB 9; Length 503;  
Best Local Similarity 67.3%; Pred. No. 0.00e+00;  
Matches 337; Conservative 93; Mismatches 68; Indels 3; Gaps 3;  
Db 1 malwrsdnlvlpptvskvlatdyvtrtniyyagssrllltvghpfaipksnsmk 88  
QY 1 MALWRSDNTVLPPTVSKVARVNTDDYVTRTSIFVHAGSSRLLTGVPYFRVPAGGNKQ 60  
Db 60 eipkvaayqyrvfvrldpdknfgldarfyndpdaerlvwactvgvgrgplvglgsh 119  
QY 61 DIPKVASAYQYRVFVRLDPDKNFGLDARFYNDPDAERLVWACTVGVGRGPLVGLSGH 120  
Db 120 plfnklldtensatnahdickqrdnsvdnyvdykqtolcilcapaigehwakgtacksrpl 179  
QY 121 PFYNKLDDESSHAATNSVSEDVRDNYVDYKQTOLCILCAPAIGEHWAKGTACKSRPL 180

Db 180 qrgdcpllelissiedgmditdfgmdftalqatkedvpidisgctckypdylkmsad 239  
QY 181 SQGDCPPELNTVLEQDMVDYIGYAMDFTSLQDTKEVPLDICISICKYDPYLQMSAD 240  
Db 240 tygnsdfflrkeqlfarhfnraghngicwgnqlfvtcvdtrstnlticstasleat 299  
QY 241 PYGDSMFFCLRRQELFAHFHFNWNRAGTMDYVQSLYIKGTCM-RASPGSCVYSPSPSGI 299  
Db 300 vtsdaqlfknkpylrragghngicwgnqlfvtcvdtrstnlticstasleat 359  
QY 300 VTSQSLFNKPYLRRAGGHNGICWGNQLFVTVVDTTSTRNLICASTQSPVPGQYD 359  
Db 360 dykfirhgeeyldqfqlckvltlpeimaylhtcmstlledwnfgltltpstaled 419  
QY 360 KFIKQYSHVEYDLQFIFQLCTITLTADVMXSIHNSNILEDNFVGPPTTSLVD 419  
Db 420 rfknsaetqcrnapvypkedpfqkfwdvdkrdfrldisqfplgrkflmglygt 479  
QY 420 RFKNSAETCQKDAAPAKENKDPYDKLFKNVNDLKEFSLDLQYPLGRKFLVQAGLRR 479  
Db 480 klgt-krlpsstssstkrkr 499  
QY 480 TIGPKRSAPSATTSSKPAKR 500  
RESULT 8  
ID VLI\_HPV30 STANDARD; PRT; 508 AA.  
AC Q02515;  
DT 01-JUL-1993 (REL. 26, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE PROBABLE L1 PROTEIN.  
GN L1.  
OS HUMAN PAPILLOMAVIRUS TYPE 30.  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.  
RN [1]  
RX MEDLINE; 94266501.  
RA DELIUS H., HOFMANN B.;  
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
RN [2]  
RX MEDLINE; 94403963.  
RA CHAN S.Y., BERNARD H.D., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;  
RL J. VIROL. 66:5714-5725(1992).  
DR EMBL; X74474; G396963; -.  
DR PIR; S36508; S36508.  
KW LATE PROTEIN.  
SQ SEQUENCE 508 AA; 56667 MW; 43C7C103 CRC32;  
Query Match 72.1%; Score 2684; DB 9; Length 508;  
Best Local Similarity 67.1%; Pred. No. 0.00e+00;  
Matches 337; Conservative 96; Mismatches 67; Indels 2; Gaps 2;  
Db 8 mavrpsaetkylpvtvskvlatdyvtrtniyyagssrllltvghpfaipksnsmk 67  
QY 1 MALWRSDNTVLPPTVSKVARVNTDDYVTRTSIFVHAGSSRLLTGVPYFRVPAGGNKQ 60  
Db 68 dpkvasayqyrvfvrldpdknfgldarfyndpdaerlvwactvgvgrgplvglgsh 127  
QY 61 DIPKVASAYQYRVFVRLDPDKNFGLDARFYNDPDAERLVWACTVGVGRGPLVGLSGH 120  
Db 128 plfnklldtensatnahdickqrdnsvdnyvdykqtolcilcapaigehwakgtacksrpl 187  
QY 121 PFYNKLDDESSHAATNSVSEDVRDNYVDYKQTOLCILCAPAIGEHWAKGTACKSRPL 180  
Db 188 agdcpllelissiedgmditdfgmdftalqatkedvpidisgctckypdylkmsad 247  
QY 181 SQGDCPPELNTVLEQDMVDYIGYAMDFTSLQDTKEVPLDICISICKYDPYLQMSAD 240  
Db 248 aygdsmyfyrreqifarfhnraghngicwgnqlfvtcvdtrstnlticstasleat 307



QY	241	pyGDSMFCLRRQLFARFWNAGTWGTVQSLYIKTGRASPGSCVYSPSPSGS	300
Db	308	tseadlnfkpywlqragghnngicwngqvftvvdttrntnmttsattqt-lst-ynsq	365
QY	301	TSDQLFNKPYWLIHKAQGHNGTCWNLQFLVTVDTTRNTTTCASTQSPVGGYDATK	360
Db	366	ikgyvzhveeyelqfvtickisalsasetmayihtmtnstillegnlgspaatalsledkyr	425
QY	361	FKQYSRHVEYDLQFIQLCTITLTADVMSYIHSMNSSILEDNWFCVPPPTSLVDIYR	420
Db	426	vykslaitcokdppaeqedpiakykfdvnlqdsfsadldqfpgkrkfmlqdvrtkps	485
QY	421	FVQSVALTCQNDRAAPENKDPYDKLFWNVDLKEKFSLOLDQIPLGKKEFVQAGURRPT	480
Db	486	tttkrsapsstststpsakrr	507
QY	481	IGPRKRSAPGATSSKPAKRRV	502

[illegible]

```

357 lkyvzhaaeyleqfjqickislaevmaylhtmnstiledwnigslppvatsledkыр 416
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 fkyosrhveydloqifqlctitladvmsyihsmnssiledwnfvgvpppttSLVDtYR 420
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 yvksaaitcoqkdpppekqdpkyskyfwevnlqnsfsadlfpqlgrkflmqvgyvtkpp 476
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 fVQSVaitCQKDAAPAEKNDPDKLAFVNVDLKEKESLDDQVPLGRKFLVQAGLRKKPT 480
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
477 vskkrasasttsapskkrkr 498
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
481 IGPKRSAPSATTSSPKAKRRV 502
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	10
ID	VLL.HPV56 STANDARD; PRT; 534 AA.
AC	P36743;
DT	01-JUN-1994 (REL. 29, CREATED)
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DI	DE PROBABLE L1 PROTEIN.
GN	L1.
OS	HUMAN PAPILLOMAVIRUS TYPE 56.
OC	VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE: 94265501.
RX	DELIOUS H.; HOFMANN B.;
RL	CURR. TOP. MICROBIOL. IMMUNOL. 186.13-31(1994).
DR	EMBL: X74483; G397059; -
DR	PIR: S36583; S36583.
KQ	LATE PROTEIN.
SQ	SEQUENCE 534 AA; 60160 MW; 6BF20335 CRC32;

```

CURR. TOP. MICROBIOL., IMMUNOL. 186:13-31(1994).
[2]
RN      RN      SEQUENCE OF 300-343 FROM N.A.
RX      RX      MEDLINE; 92407963.
RA      CHAN S.Y., BERNARD H.U., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;
J. VIROL. 66:5714-5725(1992).
DR      ENBL; X74482; G397052; -
DR      ENBL; M96298; G333194; -
DR      PIR; S36531; S36531.
KW      LATE PROTEIN.
SQ      SQ      SEQUENCE 499 AA; 55722 MW; 6B5D0108 CRC32;

Query Match          71.4%; Score 2656; DB 9; Length 499;
Best Local Similarity 66.3%; Pred. No. 0.00e+00;
Matches 333; Conservative 98; Mismatches 67; Indels 4; Gaps

Db    QY    1 mavrpsdskvylppptpvskvittdaykrttlfyhgssrrlltgphypypiskag--ka 58
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    1 MALWRPSDNVTVILPPPSVARVWVDYVTRTSIFYHAGSRLLTVGNFVRVPAGGKNQ 60

Db    Db    59 dipksafgyrvfrvlbpdpnkfglpdtnlfnpdqerlywacvgleigrqqplgvgsgh 118
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    QY    61 DIPKVSAYQYRVFRVOLPDNPDPNKGFLPDNSIYNPETQLRWACAGVEIGRQPLGVGLSG 120

Db    Db    119 plfnridtessiaqlqtapdsrdnsvdpkqtqlciigcappahehwktactsrtp 178
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    QY    121 PFYNKLDDTSSHAATSNNVEDYRDNVSDYRKQTQCIILGCAPAIGHAWAKTACKSRPL 180

Db    Db    179 tagdcppllelnspiedgdmvtgfalnfkalqeskdvpldivgstckpydkmsad 238
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    QY    181 SQGDCLPPELKNVLEGGDWDTGYGANMFSTLDQTKCEVPDLICOSICKYPDYIQMSAD 240

Db    Db    239 aygdsmwfylirreqflftrhnragvigeesipndlyiksgngrdpppsvvyatpgemi 298
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    QY    241 PYGDSMEFFCLRRELQFAHFWMNRAGTMGDIVPOSLYIKGTGMREASPGSCVSPSPGSIV 300

Db    Db    299 tseaqifnkpwylaqqghngicwnnqlfvttvdtrntnmclsatgts-mst-ynskq 356
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    QY    301 TSDQLFNKFYLHKQAQHNGICWHNQLFVTVDVTRTNLTICASTQSPPVGQYDATK 360

```

```
RESULT 11
ID VLI_HPV1 STANDARD; PRT; 501 AA.
AC P22163;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS RHESUS PAPILLOMAVIRUS TYPE 1 (HPV 1).
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91135018.
RA OSTROW R.S., LABRESE K.V., FARAS A.J.;
RL VIROLOGY 181:424-429(1991).
DR EMBL; M60184; G1019919; -.
DR PIR; H38503; P1WL1.
KW LATE PROTEIN.
SQ SEQUENCE 501 AA; 55635 MW; B0D7B721 CRC32;

Query Match 71.0%; Score 2641; DB 9; Length 501;
Best Local Similarity 67.1%; Pred. No. 0.00e+00;
Matches 338; Conservative 87; Mismatches 75; Indels 4; Gaps 4;

Db 1 msnrpsdkvlpvpsvkvstdeyartsiiyhagssrllavghpyavkkn-nkv 59
Qy 1 MALWRPSDNTVLPSPVARVNTDDYVTRTSIFYHAGSSRLLTGPNYFRVPAGGNKQ 60
Db 60 svpksvlglyrvfrvlpdpnkfgldanfydpntqrlwacilgvgevgqplvgtsgh 119
Qy 61 DIPKVSAYQYRVFVQLPDPNKFGLPDNSIYNPQRLVWACAGVEIGRQPLVGLSGH 120
Db 120 pllnkliddtengpkvaggqadrecvsnvdyktqlcmkcpvgehwknpctt-g- 177
Qy 121 PFYNKLDDESSHAATSNVSEDVNDVYKOTQOLCILGCAIPAIGEHAKGTACKSRPL 180
Db 178 aagdcplalvsnvsgdndvgygamdafnalcanksdvpidictvckypdylkmsad 237
Qy 181 SQGDCPPLKNTVLEDGDMVDYGYGAMDFSLQDTKCEVPLDICOSICKYDYLQMSAD 240
Db 238 pygdsilfvyrrgmfrhfnragtmdsvddylkagsnkvlaeshfyfptsgsmv 297
Qy 241 PYGDSMFFCLREQLFAHFHNWAGTGTVPQSLYKGTGMASPGSCVYSPSPSGSIV 300
Db 298 tsdaqlfnkpyvlqagghngicwngvflvtdttrstnmtlcatastvttyppnnes 357
Qy 301 TSDQSFNFAPYHLKAGHNGICWNNQFLVTVVDTTRSTNLTICASTQSPVPGQYDATK 360
Db 358 fkeylthveefdlqfqlckvltntevmayihmdasiledwnfnlqpppsglqdyr 417
Qy 361 FKQYSHRVEYDLQFQLCTITLTADVMSYIHSNNSILEDWNFGVPPPTSLVDYR 420
Db 418 fvtasaaltcqpppkedplaktyfvevdikefsadldqfplgrkflqagmrarpt 477
Qy 421 FVQSVAITCQDAAPAKNPDPYDKLFWNVLDKFKFSLDLDDQYPLGRKFLVQGLRRKP 480
Db 478 lrapkita-sstssssprkkrk 500
Qy 481 IGRKRSAPSATTSKPAKRVVR 504

RESULT 12
ID VLI_HPV51 STANDARD; PRT; 504 AA.
AC P26536;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 51.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.

RESULT 13
ID VLI_HPV31 STANDARD; PRT; 504 AA.
AC P17388;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 31.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89299478.
RA GOLDSBOROUGH M.D., DISILVESTRE D., TEMPLE G.F., LORINCZ A.T.;
RL VIROLOGY 171:306-311(1989).
DR EMBL; J04353; G459922; -.
DR PIR; G32444; P1WL31.
KW LATE PROTEIN.
SQ SEQUENCE 504 AA; 56352 MW; A8A44931 CRC32;

Query Match 70.5%; Score 2624; DB 9; Length 504;
Best Local Similarity 65.1%; Pred. No. 0.00e+00;

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91303675.
RA LUNGU O., CRUM C.P., SILVERSTEIN S.J.;
RL J. VIROL. 65:4216-4225(1991).
DR EMBL; M62877; -. NOT_ANNOTATED_CDS.
DR PIR; G40415; P1WL51.
KW LATE PROTEIN.
SQ SEQUENCE 504 AA; 56314 MW; EESFA804 CRC32;

Query Match 71.0%; Score 2643; DB 9; Length 504;
Best Local Similarity 64.0%; Pred. No. 0.00e+00;
Matches 323; Conservative 112; Mismatches 67; Indels 3; Gaps 3;

Db 1 malwrtndskvlpapvsvrlynteeeylirtgtyyagssrlitlghbypfpipkte-tra 59
Qy 1 MALWRPSDNTVLPSPVARVNTDDYVTRTSIFYHAGSSRLLTGPNYFRVPAGGNKQ 60
Db 60 alpkvaafgyrvfrvqlpdpnkfgldpnlpndtdrllwgcvgvevgrgqplvglsgh 119
Qy 61 DIPKVSAYQYRVFVQLPDPNKFGLPDNSIYNPQRLVWACAGVEIGRQPLVGLSGH 120
Db 120 plfnkydtenariangnaqqdvrdntsvdnkqtlclciigcappigehwigtctnptv 179
Qy 121 PFYNKLDDESSHAATSNVSEDVNDVYKOTQOLCILGCAIPAIGEHAKGTACKSRPL 180
Db 180 ppqdcpllelvsavlgdgmldtfgamdfaalqatksdvpldisqsvckypdylkmsad 239
Qy 181 SQGDCPPLKNTVLEDGDMVDYGYGAMDFSLQDTKCEVPLDICOSICKYDYLQMSAD 240
Db 240 tygnsmffhlrtreqifaryhnykllvggedipndvvyikgsgngrdpiesviysatpsgm 299
Qy 241 PYGDSMFFCLREQLFAHFHNWAGTGTVPQSLYKGTGM-RASPGSCVYSPSPSGSI 299
Db 300 itsdqlfnkpyvlhraqghngicwngqlfctvtdttrstnltstat-aavspftps 358
Qy 300 VTSQSLFNKPYHLKAGHNGICWNNQFLVTVVDTTRSTNLTICASTQSPVPGQYDAT 359
Db 359 nfkylrthveefdlqfqlckvltntevmayihmdotilegwnfgltlppasleday 418
Qy 360 FKQYSHRVEYDLQFQLCTITLTADVMSYIHSNNSILEDWNFGVPPPTSLVDYR 419
Db 419 fvtasaaltcqpppkedplaktyfvevdikefsadldqfalgkrkflvgvqrkp 478
Qy 420 FVQSVAITCQDAAPAKNPDPYDKLFWNVLDKFKFSLDLDDQYPLGRKFLVQGLRRKP 479
Db 479 rpglkrpassasasasasakkrvk 503
Qy 480 TIGPKRSAPSATTSKPAKRVVR 504
```

Matches 329; Conservative 102; Mismatches 71; Indels 3; Gaps 3;

```

Db 1 malwrsneatvylppvsvskvstdevtrtniyyhagsarlllvghpyvaikkqsnk1 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MALWRSDNTVLPSPSVARVNTDDYVTRTSIFYHAGSRLLTVGNFVRYPAGGNGK 59
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 ivpksvglqyrvfrvlpdpnkfgfdtsfynpctqrlwacvlgvrgqplgvlgsh 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 60 QDPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPOTQRLVWACAGVEIGRQPLGVLSG 119
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 hpllnklddtenanryagpgtdnrecismdykqtlclllgckpplgchwkgkspcnaa 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 120 HFYKLDDESSHAATSNVSEDVRNVSDYKQTLQILGCAPAIGERHAWKGTACKSRP 179
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 itpgdcpllelnknsiqdgmvdgtfgamdfatqtkenavpldicnsickypdyikmva 240
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 180 LSGDCPPELKNVLELGDGMVDGTGYGAMDFSLQTKCEVPLDICQICKYDPYLQMSA 239
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 epvgdltfllylrreqmfvrhfrstgvesvptdllykgsstatlaustyftpsgm 300
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 240 DPGDSMFFCLREQLFARHFWNRAGTMDTPVQSLYIKGTMRASPGSCVSPSGSI 299
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 301 vrsdaqifnkyvmqraqhngicwnglfvtvdttrstmsvcaaiansdtt-fks 359
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 300 VTSDQLFNKPYLHLKRAQGHNGICWHNQLFVTVVDTTRSTNLITICASTQSPVPGQYDAT 359
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 nkeylrhgeeydlqfqlckitladmtyihmmpalledwnfglttppsgsldety 419
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 360 KFOYSRHVEYDLQFQLCTITLADVMYSYHSMNSLIEDWNFGVPPPTSLVDYR 419
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 rfvtsqalcqtapdkpkedpfkdyfvevnlkefsadldqfplgrkflqagyrarp 479
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 420 RFVQSVAITCQKDAAPAEKNDPYDKLFWNVLKKEFSLDLDQYPLGRKFLVQAGLRKP 479
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 480 kfkagkraspsattt-pakrktk 503
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 480 TIGPKRSAPSATTSSKPAKRVVR 504
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 14
ID VLI.HPV35 STANDARD; PRT; 502 AA.
AC P2732;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 35.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 94265501.
RA DELIUS H., HOPMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92124753.
RA MARICH J.E., PONTSLER A.V., RICE S.M., MCGRAW K.A., DUBENSKY T.W.;
RL VIROLOGY 186:770-776(1992).
DR EMBL; X74477; G397004; -.
DR EMBL; M74117; G333058; -.
DR PIR; G40824; P1WL35.
DR PIR; S36526; S36526.
KW LATE PROTEIN.
FT CONFLICT 94 95 AS -> CL (IN REF. 2).
FT CONFLICT 132 132 S -> L (IN REF. 2).
FT CONFLICT 140 140 G -> GNSG (IN REF. 2).
FT CONFLICT 482 482 K -> R (IN REF. 2).
SQ SEQUENCE 502 AA; 56148 MW; 0EF602B1 CRC32;

```

Query Match 69.9%; Score 2600; DB 9; Length 502;  
 Best Local Similarity 65.1%; Pred. No. 0.00e+00;  
 Matches 329; Conservative 95; Mismatches 78; Indels 3; Gaps 3;

```

Db 1 malwrsneatvylppvsvskvstdevtrtniyyhagsarlllvghpyvaikkqsnk1 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MALWRSDNTVLPSPSVARVNTDDYVTRTSIFYHAGSRLLTVGNFVRYPAGGNGK 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 avpkvsglyrvfrvlpdpnkfgfdtsfydpasqrlwactvgvrgqplgvlgsh 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPOTQRLVWACAGVEIGRQPLGVLSG 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 pllnklddtenanryagpgtdnrecismdykqtlclllgckpplgchwkgktpcnaqy 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 PFYKLDDESSHAATSNVSEDVRNVSDYKQTLQILGCAPAIGERHAWKGTACKSRPL 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 kagecppllelnvlgdgmvdgtfgamdfatqtkenavpldicnsickypdyikmva 240
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 SQGDCPPELKNVLELGDGMVDGTGYGAMDFSLQTKCEVPLDICQICKYDPYLQMSAD 240
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 pygdmflfyllrreqmfvrhfragvtgvetpadlykgtt-tlpsts-yftpsgm 298
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 PYGDSMFFCLREQLFARHFWNRAGTMDTPVQSLYIKGTMRASPGSCVSPSGSI 300
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 299 tsdaqifnkyvmqraqhngicwnglfvtvdttrstmsvcaavss-dstykn 357
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 TSDQLFNKPYLHLKRAQGHNGICWHNQLFVTVVDTTRSTNLITICASTQSPVPGQYDAT 360
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 358 fkeylrhgeeydlqfqlckitladmtyihmmpalledwnfglttppsgtledtyr 417
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 KFOYSRHVEYDLQFQLCTITLADVMYSYHSMNSLIEDWNFGVPPPTSLVDYR 420
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 418 ytsqavtcqkpsapkdpdknkytfevdlkefsadldqfplgrkflqaglkarpn 477
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 FVQSVAITCQKDAAPAEKNDPYDKLFWNVLKKEFSLDLDQYPLGRKFLVQAGLRKPT 480
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 478 flgrkraapastekstkrkvks 502
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 481 IGPKRSAPSATTSSKPAKRVVR 505
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 15
ID VLI.HPV2A STANDARD; PRT; 510 AA.
AC P25486;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 2A.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 91188699.
RA HIRSCH-BEHNAM A., DELIUS H., DE VILLIERS E.M.;
RL VIRUS RES. 18:81-98(1990).
DR EMBL; X55964; -; NOT_ANNOTATED_CDS.
DR PIR; S15620; S15620.
KW LATE PROTEIN.
SQ SEQUENCE 510 AA; 57193 MW; EB7FCF70 CRC32;

```

Query Match 69.7%; Score 2592; DB 9; Length 510;  
 Best Local Similarity 64.1%; Pred. No. 0.00e+00;  
 Matches 323; Conservative 111; Mismatches 59; Indels 11; Gaps 9;

```

Db 17 malwrsneatvylppvsvskvstdevtrtniyyhagsarlllvghpyvaikkna-nkv 75
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 MALWRSDNTVLPSPSVARVNTDDYVTRTSIFYHAGSRLLTVGNFVRYPAGGNGK 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 76 avpkvsglyrvfrvlpdpnkfgfdadlypdtqrlwacvlgvrgqplgvlgsh 135
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPOTQRLVWACAGVEIGRQPLGVLSG 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 136 pynrlddtenah--tpdtaddgrenismdykqtlclllgckpplgchwskgttcngss- 192
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 PFYKLDDESSHAATSNVSEDVRNVSDYKQTLQILGCAPAIGERHAWKGTACKSRPL 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

Db	193	aagdcpplqfntttledmvetgafalafalqnsksdvpldictnckypdylnmae	252
Qy	181	SGQOCPPPEKUNVLEDGMDVDTGAMDFSLQDTKCEVPLDQCICKPTQLQMSAD	240
Db	253	pygdmffslrreqmfthffnlggkmgdtlpdelylksstv-ptpghvystptpgsmv	311
Qy	241	PIGDSMEFLCRRRQQLFAHFWRNAGTMTGTPQSLYIKGTGMRASPCGVSPSPSGSIV	300
Db	312	sseoglfnkpylrragqhnmgcwgvrfltwdvttrstnvslica-teas-dtnykatn	369
Qy	301	YSDSQLFNKPYLHLKAAQGNNGICWNLQFLVTVDTFRSTNLTCASQSPVQGDYATK	360
Db	370	fkexylrhmeydlqficklxtlptelmaylthmdpqllcdwnfgvppppaslsqdyr	429
Qy	361	FQQSRVIEEYDLQIFOLCTITLTADVMSYIHSMNSSILEDWNFGVPPPTTSLVDTYR	420
Db	430	ylqsqaitcktpbpktpdtpyalstfwdvdisesfmdldqfgrkflqlrga--mpt	487
Qy	421	FVQSVATTCOKAAPENKNDPDKLAFNVNVLKXFLSDLDQYPLGRKFLVQAGLRKKPT	480
Db	488	vs-rkraavgttptpt-skrkrvr	509
Qy	481	IGPKRSGAPSAITSSKPAKRVRR	504

Search completed: Wed Mar 26 08:27:53 1997  
Job time : 104 secs.











































\*\*\*\*\*

W P S R L H

(TW)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Mar 26 08:31:24 1997; MasPar time 139.35 Seconds  
1216.007 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-409-122-1

Description: (1-1524) from US08409122.seq

Perfect Score: 7620

N.A. Sequence: 1 ATGGCTTTGTGGCGGCTAG.....GTGTAGTGCACGAGGAATAA 1524

Comp: TACCGAATACCGCGGATC.....CACATGACGGCTCTTCATT

Scoring table: TABLE bktranslate2

Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 52205 seqs, 55594155 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot33

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10

Statistics: Mean 76.762; Variance 127.210; scale 0.603

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	5890	77.3	568	9	VLI_HPV18 PROBABLE L1 PROTEIN.	0.00e+00
2	4888	64.1	539	9	VLI_HPV45 PROBABLE L1 PROTEIN.	0.00e+00
3	4551	59.7	505	9	VLI_HPVME PROBABLE L1 PROTEIN.	0.00e+00
4	4449	58.4	505	9	VLI_HPV39 PROBABLE L1 PROTEIN.	0.00e+00
5	3797	49.8	531	9	VLI_HPV10 PROBABLE L1 PROTEIN.	0.00e+00
6	3743	49.1	508	9	VLI_HPV10 PROBABLE L1 PROTEIN.	0.00e+00
7	3714	48.7	503	9	VLI_HPV26 PROBABLE L1 PROTEIN.	0.00e+00
8	3530	46.3	504	9	VLI_HPV31 PROBABLE L1 PROTEIN.	0.00e+00
9	3522	46.2	532	9	VLI_HPV03 PROBABLE L1 PROTEIN.	0.00e+00
10	3504	46.0	501	9	VLI_HPV01 PROBABLE L1 PROTEIN.	0.00e+00
11	3505	46.0	504	9	VLI_HPV51 PROBABLE L1 PROTEIN.	0.00e+00
12	3489	45.8	499	9	VLI_HPV53 PROBABLE L1 PROTEIN.	0.00e+00
13	3493	45.8	534	9	VLI_HPV56 PROBABLE L1 PROTEIN.	0.00e+00
14	3446	45.2	531	9	VLI_HPV16 PROBABLE L1 PROTEIN.	0.00e+00
15	3415	44.8	502	9	VLI_HPV35 PROBABLE L1 PROTEIN.	0.00e+00
16	3400	44.6	505	9	VLI_HPV40 PROBABLE L1 PROTEIN.	0.00e+00
17	3389	44.5	505	9	VLI_HPV07 PROBABLE L1 PROTEIN.	0.00e+00
18	3257	42.7	499	9	VLI_HPV33 PROBABLE L1 PROTEIN.	0.00e+00

19	3248	42.6	524	9	VLI_HPV58 PROBABLE L1 PROTEIN.	0.00e+00
20	3205	42.1	502	9	VLI_PCPV1 PROBABLE L1 PROTEIN.	0.00e+00
21	3176	41.7	594	9	VLI_HPV27 PROBABLE L1 PROTEIN.	0.00e+00
22	3152	41.4	528	9	VLI_HPV34 PROBABLE L1 PROTEIN.	0.00e+00
23	3084	40.5	529	9	VLI_HPV52 PROBABLE L1 PROTEIN.	0.00e+00
24	3067	40.2	510	9	VLI_HPV57 PROBABLE L1 PROTEIN.	0.00e+00
25	3044	39.9	501	9	VLI_HPV6B PROBABLE L1 PROTEIN.	0.00e+00
26	3030	39.8	501	9	VLI_HPV11 PROBABLE L1 PROTEIN.	0.00e+00
27	3023	39.7	510	9	VLI_HPV2A PROBABLE L1 PROTEIN.	0.00e+00
28	3020	39.4	499	9	VLI_HPV13 PROBABLE L1 PROTEIN.	0.00e+00
29	2989	38.2	502	9	VLI_HPV42 PROBABLE L1 PROTEIN.	0.00e+00
30	2928	38.4	503	9	VLI_HPV32 PROBABLE L1 PROTEIN.	0.00e+00
31	2133	28.0	501	9	VLI_PAPVE PROBABLE L1 PROTEIN.	0.00e+00
32	2111	27.7	507	9	VLI_HPV63 PROBABLE L1 PROTEIN.	0.00e+00
33	1999	26.2	505	9	VLI_CRPVK PROBABLE L1 PROTEIN.	0.00e+00
34	1967	25.8	495	9	VLI_BPVI PROBABLE L1 PROTEIN.	0.00e+00
35	1965	25.8	497	9	VLI_BPVI PROBABLE L1 PROTEIN.	0.00e+00
36	1889	24.8	513	9	VLI_PAPVD PROBABLE L1 PROTEIN.	0.00e+00
37	1884	24.7	507	9	VLI_HPV17 PROBABLE L1 PROTEIN.	0.00e+00
38	1873	24.6	507	9	VLI_HPV09 L1 PROTEIN.	0.00e+00
39	1852	24.3	516	9	VLI_HPV65 PROBABLE L1 PROTEIN.	0.00e+00
40	1834	24.1	516	9	VLI_HPV04 PROBABLE L1 PROTEIN.	0.00e+00
41	1815	23.8	507	9	VLI_HPV15 PROBABLE L1 PROTEIN.	0.00e+00
42	1809	23.7	508	9	VLI_HPV1A PROBABLE L1 PROTEIN.	0.00e+00
43	1768	23.2	509	9	VLI_HPV49 PROBABLE L1 PROTEIN.	0.00e+00
44	1660	21.8	506	9	VLI_BPVI PROBABLE L1 PROTEIN.	0.00e+00
45	1585	20.8	518	9	VLI_HPV14 PROBABLE L1 PROTEIN.	0.00e+00

ALIGNMENTS

RESULT	ID	VLI_HPV18	STANDARD;	PRT;	568 AA.
AC	P06794;				
DT	01-JAN-1988 (REL. 06, CREATED)				
DT	01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)				
DT	01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)				
DE	PROBABLE L1 PROTEIN.				
GN	L1				
OS	HUMAN PAPILLOMAVIRUS TYPE 18.				
OC	VIROIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.				
RN	[1]				
RX	SEQUENCE FROM N.A.				
RA	COLE S.T., DANOS O.;				
RA	J. MOL. BIOL. 193:599-608(1987).				
RN	[2]				
RN	SEQUENCE OF 363-406 FROM N.A.				
RA	CHAN S.Y., BERNARD H.U., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;				
RA	MEDLINE; 92407953.				
RL	J. VIROL. 66:5714-5725(1992).				
DR	EMBL; X05015; G60983; -				
DR	EMBL; M96287; G333172; -				
DR	EMBL; A06329; G413678; -				
DR	PIR; A26251; P1WL18.				
KW	LATE PROTEIN.				
FT	CONFLICT 384 384 V -> I (IN REF. 2).				
SQ	SEQUENCE 568 AA; 63623 MW; 8D0D12FC CRC32;				
Query Match 77.3%; Score 5890; DB 9; Length 568;					
Best Local Similarity 57.7%; Pred. No. 0.00e+00;					
Matches 877; Conservative 354; Mismatches 290; Indels 0; Gaps 0;					
Db	m a l w r p s d n t v y l p p s v a r				
Dt	atggcnyntgtgmcncnwgaaayacnqntatytncncncncnsvngtngcmgn 243				
Qy	1 ATGGCTTTGTGGCGGCTAGACATACCGTATACCTCCACCTCTCTGTGGAAGA 60				
Qt	M A L W R P S D N T V Y L P P S V A R				
Db	v y n t d d y v t p t s i f y h a g s s				
Dt	gtgtatayacngaygtaygtatnncnncnncnncnncnncnncnncnncnncn 303				
Qy	61 GTGTATATACATCAIGATITGIGTACCTCCACACACATATTTATCATCTGCGACCTCT 120				





















```

QY 1441 ATAGCCCTCTAAGCTTCGTCCTCACTGCTTAAACGCGCAAGCGT 1500
Qt I G P R K R S A P S A T T S S K P A K R

Db
Dt 1516 aarnng 1520
QY 1501 GTGCG 1505
Qt V

RESULT 7
ENTRY S36549 #type complete
TITLE late protein - human papillomavirus type 26
ORGANISM #formal_name human papillomavirus type 26
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36549
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36549
#status Preliminary
#molecule_type DNA
#residues 1-503 #label DEL
#cross-references EMBL:X74472
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 503 #molecular-weight 56227 #checksum 9874

Query Match 48.7%; Score 3714; DB 6; Length 503;
Best Local Similarity 48.9%; Pred. No. 0.00e+00;
Matches 704; Conservative 296; Mismatches 435; Indels 6; Gaps 5;

Db m a l w r t s d s k v y l p p t p v s r
Dt 1 atgcnctntgmgnaacnswgaysnaargntntaytncnccnaccnccngtnswmgn 60
QY 1 ATGGCTTTGTGCGCGCTAGTGACAATACCGTATACCTTCCACCTCTCTGTCGCAAGA 60
Qt M A L W R P S D N T V Y L P P P S V A R

v v n t d e y v t r t g i y y a g s s
Dt 61 gtngnaayacngaygartaygtacnmgnaacngnathaytaygngngnswm 120
QY 61 GTTGTAACTAGTATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qt V V N T D D Y V T R T S I F Y H A G S S

r l l t l g h p y f s i p k t g q k a e
Dt 121 mgnytynacnngnccaycncctaytysnathcna-aracng-gncaraargng 178
QY 121 AGATTATTAACTGTGGTAATCCATATTTTAGGTTCCTGCGAGTGCGCAATAAGCAG 180
Qt R L L T V G S A Y Q Y R V F R V P A Q * A G

i p k v s a y q y r v f r v h l p d p
Db 179 ar-athcnaargntnngcngcaycncctaytysnathcna-aracng-gncaraargng 237
Dt 179 ar-athcnaargntnngcngcaycncctaytysnathcna-aracng-gncaraargng 237
QY 181 GATATTCCTAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Qt I P K V S A Y Q Y R V F R V P A Q * A G

n k f g l p d p q l y n p d t e r l v w
Db 238 aayaarttngnngcngcaycncctaytysnathcna-aracng-gncaraargng 297
Dt 238 aayaarttngnngcngcaycncctaytysnathcna-aracng-gncaraargng 297
QY 241 AATAAATTTGCTTACCTGATAAGTATTTATATCTGTAACACACACACACACACAC 300
Qt N K F G L P D N S I Y N P E T Q R L V W

a c v g v e v r g q p l g i g l s w g h
Db 298 gcntgngngngngngngngngngngngngngngngngngngngngngngngngngng 357
Dt 298 gcntgngngngngngngngngngngngngngngngngngngngngngngngngngng 357
QY 301 GCGTGTGCTGAGTGAATTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qt A C A G V E I G R G Q P L P L S G H

p l f n k l d d t e n s h l a t v n a d
Db 358 ccnyntntyaayaryngaygacngaraaywncayyngnaccngntnaaygcnagay 417
Dt 358 ccnyntntyaayaryngaygacngaraaywncayyngnaccngntnaaygcnagay 417

```

```

QY 361 CCATTTTATAATAATTAGATGACACTGAAGTTCCCATGCGCTAGCTTAAGTTTCT 420
Qt P F Y N K L D D T E S S H A A T S N V S

t d n r d n v s v d n k g t q l c i i g
Db 418 acngayayngngayaygntnswngtngayaayaaaracnaccnaryntngyathngn 477
Dt 418 acngayayngngayaygntnswngtngayaayaaaracnaccnaryntngyathngn 477
QY 421 GAGGACGTTAGGACAATGCTGCTAGATATAACAGACACAGATTTATTTGGGCG 480
Qt E D V R D N V S V D Y K Q T Q L C I L G

c t p p l g e h w g i g t i c k n t q t
Db 478 tgyacnccnccnytnngngngngngngngngngngngngngngngngngngngngng 537
Dt 478 tgyacnccnccnytnngngngngngngngngngngngngngngngngngngngngng 537
QY 481 TGTGCTGCTGCTATTTGGGACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qt C A P A I G E H W A K G T A C K S R P L

q r g d c p p l e l i s s i i e d g d m
Db 538 carngngngaytgcncnccnytnngngngngngngngngngngngngngngngngng 597
Dt 538 carngngngaytgcncnccnytnngngngngngngngngngngngngngngngngng 597
QY 541 TCACAGGCGGATTCGCCCTTTAGAACTTAAGAACACAGATTTTGAAGATGGTATG 600
Qt S Q G D C P P L E L K N T V L E D G D M

i d t g f g a m d f t a l q a t k s d v
Db 598 athayacngntngngngngngngngngngngngngngngngngngngngngngng 657
Dt 598 athayacngntngngngngngngngngngngngngngngngngngngngngngng 657
QY 601 GTAGATAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qt V D T G Y G A M D F S T L Q D T K C E V

p i d i s q s t c k y p d y l k m s a d
Db 658 ccnathayahwncnccnccnytnngngngngngngngngngngngngngngngngng 717
Dt 658 ccnathayahwncnccnccnytnngngngngngngngngngngngngngngngngng 717
QY 661 CCATTTGATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Qt P L D I C Q S I C K Y P D Y L Q M S A D

t y g n s m f f l r x e q l f a r h f
Db 718 acncaayngnaaywsnattgttytntngngngngngngngngngngngngngngngng 777
Dt 718 acncaayngnaaywsnattgttytntngngngngngngngngngngngngngngngng 777
QY 721 CCTTATGGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Qt P Y G D S M F F C L R R E Q L F A R H F

y n k a g a v g d a i p t t l y i k g a
Db 778 tayaaayargngngngngngngngngngngngngngngngngngngngngngngng 837
Dt 778 tayaaayargngngngngngngngngngngngngngngngngngngngngngngng 837
QY 781 TGGAAATAGGAGGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Qt W N R A G T M G D T V P Q S L Y I K G T

e s g r e p p t s s i y s a t p s g s m
Db 838 garwngngngngngngngngngngngngngngngngngngngngngngngngngng 897
Dt 838 garwngngngngngngngngngngngngngngngngngngngngngngngngngng 897
QY 841 G-GTATG-CGTGCTTACCTGCGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
Qt X Y X R A S P G S C V Y S P S P S G S I

v t s d a q l f n k p y w l q r a q g h
Db 898 gtnacnswngngngngngngngngngngngngngngngngngngngngngngngng 957
Dt 898 gtnacnswngngngngngngngngngngngngngngngngngngngngngngngng 957
QY 898 GTTACCTCTGACTCCAGTTGTTTAAATAACCATATGTTTACATAAGGACAGGTCAT 957
Qt V T S D S Q L F N K P Y W L H K A Q G H

n n g i c w g n q l f v t c v d t t r s
Db 958 aayaayngnattgttggngnaaycaryntngntngntngntngntngntngntngntng 1017
Dt 958 aayaayngnattgttggngnaaycaryntngntngntngntngntngntngntngntng 1017
QY 958 AACAAATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
Qt N N G I C W H N Q L F V T V V D T T R S

t n l t i s t l s a a s a s t p f k p s
Db 1018 acnaaytynacnathwncnccnccnytnngngngngngngngngngngngngngngng 1077
Dt 1018 acnaaytynacnathwncnccnccnytnngngngngngngngngngngngngngngng 1077
QY 1018 ACCAATTAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
Qt T N L T I C A S T Q S P V P G Q Y D A T

d y k q f i r h g e e l q f i q l
Db 1078 gaytayaarcarttghmgncayngngngngngngngngngngngngngngngngngng 1137
Dt 1078 gaytayaarcarttghmgncayngngngngngngngngngngngngngngngngngng 1137
QY 1078 AAATTTAAGCAGTATAGCAGACATGTTGAAGAATATGATTTGCAAGTTATTTTCA 1137
Qt

```

Dt	61	gtngrnswsnaacngaygartaygtatnccnmgnaacnaayathaytaycaygcnngnwnsgn	120
Qy	61	GTGTPAATAATCATGATTATGACTCGCAACAGCATATTTATCATGTGTGGCAGCTCT	120
Qt		V V V N T D D Y V T R T S I F Y H A G S S	
Dd		r l l t t v g h p c y t a y w s n a t h p k k s d n p c n a a r 180	
Dt	121	mgnynlncnagncaycncntaytaywsnathpckkssdnpcnaa180	
Dt	121	AGATTATTAACCTGTGGTAATCCATATTTTAGGTTCCTCGAGGTGTGGCAATAAGCAG	180
Qt		R L L T V G N P Y F R V P A G G G N K Q	
Dd		i v v p k v s g l q y r v f r v r i p k s d n p c n a a r 240	
Dt	181	athgngnccncaargntgcyataymngntgntymngnmgnytnccngay 240	
Qy	181	G-A-TA-TTCCTAAGCTTCGCATACCAATATAGATGATTTTCGGTGCAGTATCCCTGAC	237
Qt		X x x p k v s a y q y r v f r v q l p d	
Dd		p n k f g f d t s f y n p e t q r l v	
Dt	241	concaeyartgntgcnaycnwstiyaycncnargncarmnngntn 300	
Qy	238	CCAAATATTTGTTTACCTGATATGATTAATATTCCTGAACACACCTGTATG 297	
Qt		P N K F L P D N S I Y N P E T Q R L V	
Dd		w a c v g l e v g r g q p l g v v g l s g	
Dt	301	tggnngyngntgngngngncarcncnytnngntngnathwngn 360	
Qy	298	TGGGCTGTGGTGGAGAAATGGCGGTGACGCTTTAGGTGTGGCCATTAGTGG 357	
Qt		W A C A G E I G R G Q P L G V G L S G	
Dd		h p l l n k f d t e n s n r y a g g p	
Dt	361	caycnytnaayarttygayaacngaraaywsnathpckkssdnpcnaa420	
Qy	358	CATCCATTTAATAAATAGATGACACTGAAAGTTCOCATGCCGTCTAATGTT 417	
Qt		H P F V N K L D D T E S S H A A T G T S N V	
Dd		g t d n r e c i s m d y k q t q l c l l	
Dt	421	ggnacngayaamngartgyathwsnatgyatyaaracnarcnaryntgnytn 480	
Qy	418	TCTGAGACGTTAGGCAACTGTCTGTAGATTATAACAGCACACAGATTATGTTG 477	
Qt		S E D V R D N V S V D Y K Q T Q L C L L	
Dd		g c k p p i g e h w g k g s p c l n a	
Dt	481	gntgyaarcncnathbgngcaytgggngnaayggncnccntgwnaaycng 540	
Qy	478	GGCTGTGCCCTGCTATTGGGACACTGGCTTAAAGGACACTGTGTAAATCGCTCT 537	
Qt		G C A P A I G E H W A K G T A C K S R P	
Dd		i t p g d c d p l e l k n s v i q d g d	
Dt	541	athacnccngnaygyccncaytgartyaarayaayngnathargayngn 600	
Qy	538	TATCACAGGCGCATGGCCCTTTAGAACCTTAAGAACACAGTTTGGAGAGATGTGAI 597	
Qt		L S Q G D C P P L E L K N T Y L E D T K S N	
Dd		m v d t g f g a m d f t a l q d t k s n	
Dt	601	atgngnngnatgngngnatgyatyaacngnytncaargayacnaaraya 660	
Qy	598	ATGGTAGATATGGATATGGTCCCATGGACTTTAGTACATTGCAAGATATAATATGAG 657	
Qt		M V D T G Y G A M D F S T L Q D T K C E	
Dd		v p l d i c n s i c k y p d y l k m v a	
Dt	661	gtnccnyngayathgyaaywsnathgyaartaycncngaytaynaaragtn 720	
Qy	658	GTACCATGGATATTTGAGTCTATTGATAATATTCCTGATTTACAAATGTCTGCA 717	
Qt		V P L D I C Q S I C K Y P D Y L Q M S A	
Dd		g p y g d t l f f y l r e q m f v r h	
Dt	721	garcentyngngaycnytnytyayytumngngngaracaragtvgtnmncay 780	
Qy	718	GATCCTATGGGATTCCTATGTTTTTGGTTTGTACGACGTGAGACGTGTGTAGCAT 777	
Qt		D P Y G D S M F F C L R R E Q L F A R H	
Dd		f y n r s g t v g e s v p t d l y i k g a	
Dt	781	tytytyaayngnswngnccngnngnccnagngnccnagngnytnathargn 840	





181 GATATCTCAAGTTCGACCAATACCAATAGATATTTCGGTGCAGTACCTGACCCA 240  
 I P K V S A Y Q Y R V F R V Q L P P  
 238 aayaaaygnnathgtytggaayaaaycaryntntthachenttygtngayachnagmwn 1017  
 958 AACATGATGATCTGCTGGCAATCAATATTTATTTGCTAGCTGGTAGACCACTGAT 1017  
 241 AATAAATTTGGTTTACCTGATAATAGTATTATAATCTGAAACACACAGTTTGTGG 300  
 N K F G L P D N S I Y N P E T Q R L V W  
 298 gnttygtng 357  
 301 GCCTGCTGAGGAGGAAATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
 A C A G V E I G R G Q P L G V G L S G H  
 358 cnynttyaayaaay 417  
 361 CCATTTTATAATAAATAGATGACACTGAAAGTTCCTGCTGCTGCTGCTGCTGCTGCT 420  
 P F Y N K L D D T E S S H A A T S N V S  
 418 carygtng 477  
 421 GAGGACGTTAGGCAATGCTGCTGATATATAGCAGACAGTATGATGATGATGATG 480  
 E D V R D N V S V D Y K Q T Q L C I L G  
 478 cnyng 537  
 481 TGTCGCCCTGATTTGGGGAACACTGGCTAAAGCACTGCTTGTAAATCGCGCTTTA 540  
 C A P A I G E H W A K G T A C K S R P L  
 538 cncng 597  
 541 TCACAGGCGATGTCGCCCTTTAGAACATTAAGACACAGTGTGGAAGATGGTGATG 600  
 S Q G D C P P L E L K N T V L E D G D M  
 598 athgayaeng 657  
 601 GTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 V D T G Y G A M D F S T L Q D T K C E V  
 658 cnyntngayathswncarwngntngayayayayayayayayayayayayayayay 717  
 661 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 P L D I C Q S I C K Y P D Y L Q M S A D  
 718 acntayng 777  
 721 CCTATGGGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 P Y G D S M F F C L R R E Q L F A R H F  
 778 tayaay 837  
 781 TGGATAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 W N R A G G T M G D T V P Q S L Y I K G T  
 838 gng 897  
 841 GATA-TG-CGTGCTACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897  
 G X R A S P G S C V Y S P S P S G S I  
 898 ithswng 957  
 athcnwng

898 GTTACCTCTGACTCCAGTCTGTTTAAATACCATATTTGTTACATAAGGCACAGGTGAT 957  
 V T S D S Q L F N K P Y W L H K A Q G H  
 958 aayaaaygnnathgtytggaayaaaycaryntntthachenttygtngayachnagmwn 1017  
 958 AACATGATGATCTGCTGGCAATCAATATTTATTTGCTAGCTGGTAGACCACTGAT 1017  
 N N G I C W H N Q L F V T V V D T T R S  
 1018 acnaayntnacnathswncngnagcn--g-cngcngntnswncnagnttyacnncnwn 1074  
 1018 ACCAATTTAAACATATGCTTACACAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1077  
 T N L T I C A S T V P V P G Q Y D A T  
 1075 aaytyaarcaryatvathngncayggngargartaygaryntncarttyaththycaryn 1134  
 1078 K F K Q Y S R H V E E Y D L Q F I F Q L  
 1135 tgyaarcachnynacnagngngngngngngngngngngngngngngngngngngng 1194  
 1138 TGTATCTATCTTAACTGACAGTGTATGCTTATCTATCTATCTATCTATCTATCTAT 1197  
 C T I T L T A D V M S Y I H S M N S S I  
 1195 ytnarcaryatvathgngngngngngngngngngngngngngngngngngngng 1254  
 1198 TTAGAGGATGAACTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1257  
 L E D W N F G V P P P P T T S L V D T Y  
 1255 rfvyrnanaataatstscqkdtpdpqakp 1314  
 1258 CGTTTGTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317  
 R F V Q S V A I T C Q K D A A P A E N K  
 1315 gaycnyngncnartayaaarttyggayngngngngngngngngngngngngngng 1374  
 1318 GATCCCTATGATAAGTTAAAGTTTGGATTTGATTTGATTTGATTTGATTTGATTTG 1377  
 D P Y D K L K F W N V D L K K E K F S L D  
 1375 ytnaycarytng 1434  
 1378 TTAGATCAATATCCCTTGGAGCTAAATTTTGGTTCAGCTGCTGCTGCTGCTGCTGCT 1437  
 L D Q Y P L G R K F L V Q A G L R R K P  
 1435 mngcngngnytnaarmngngngngngngngngngngngngngngngngngngng 1494  
 1438 ACCATGAGCCCTGCTAAACGTTCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 1497  
 T I G P R K R S A P S A T T S S K P A K  
 1495 mngnaarmngnt 1505  
 1498 CGTGTGCTGT 1508  
 R V R  
 RESULT 11  
 ENTRY #type complete  
 TITLE late protein - human papillomavirus type 53  
 ORGANISM formal name human papillomavirus type 53  
 DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 12-Apr-1995  
 ACCESSIONS S36531  
 REFERENCE S36469  
 #authors Dellius, H.; Hofmann, B.



```

Db 1309 cnyntnswnaartayarttgggargtnaaytncaraaywsnttywsngcngayvtn 1368
Dt 1309 cnyntnswnaartayarttgggargtnaaytncaraaywsnttywsngcngayvtn 1368
QY 1321 CCTATGATAAGTAAAGTTTGGAATGGATTGAAGGARAAGTTTCTTTGGACTTA 1380
Qt 1321 CCTATGATAAGTAAAGTTTGGAATGGATTGAAGGARAAGTTTCTTTGGACTTA 1380
Db 1369 gaycarttccnyntngnmgnaarttyttnatcargtngngngnmgnaacnncn 1428
Dt 1369 gaycarttccnyntngnmgnaarttyttnatcargtngngngnmgnaacnncn 1428
QY 1381 GATCAATATCCCTTCGAGCTAAATTTTGGTTGAGCTGGATGGTGGCAGCCACC 1440
Qt 1381 GATCAATATCCCTTCGAGCTAAATTTTGGTTGAGCTGGATGGTGGCAGCCACC 1440
Db 1429 gtnswnsnaaraarngnswngnswnaacnswnaacnswngnswnaacnswnaarmgn 1488
Dt 1429 gtnswnsnaaraarngnswngnswnaacnswnaacnswngnswnaacnswnaarmgn 1488
QY 1441 ATAGGCCCTCGTAACGTTCTGCTCCATCTGCCACACGCTTCTTAACCTGCCAAGCGT 1500
Qt 1441 ATAGGCCCTCGTAACGTTCTGCTCCATCTGCCACACGCTTCTTAACCTGCCAAGCGT 1500
Db 1489 aarmg 1493
Dt 1489 aarmg 1493
QY 1501 GTGCG 1505
Qt 1501 GTGCG 1505

RESULT 12
ENTRY S36583 #type complete
TITLE late protein - human papillomavirus type 56
ORGANISM #formal_name human papillomavirus type 56
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 12-Apr-1995
ACCESSIONS S36583
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36583
#status preliminary
#molecule_type DNA
#residues 1-534 #label DEL
#cross-references EMBL:X74483
CLASSIFICATION #superfamily papillomavirus I1 protein
SUMMARY #length 534 #molecular_weight 60160 #checksum 1336
Query Match 45.8%; Score 3493; DB 6; Length 534;
Best Local Similarity 47.7%; Pred. No. 0.00e+00;
Matches 694; Conservative 292; Mismatches 460; Indels 9; Gaps 5;

Db 106 atggcnaatgngnswngaraayarttntaytntccnccnccngnswnaar 165
Dt 106 atggcnaatgngnswngaraayarttntaytntccnccnccngnswnaar 165
QY 1 ATGCGCTTGTGGGGCTAGTGACATACCGTATACCTCCACCTCCTTCTGTGGCAAGA 60
Qt 1 ATGCGCTTGTGGGGCTAGTGACATACCGTATACCTCCACCTCCTTCTGTGGCAAGA 60
Db 166 gtngtngnacngayntaytngnaarmgnacnswnaathtytaycayngcngnswnsn 225
Dt 166 gtngtngnacngayntaytngnaarmgnacnswnaathtytaycayngcngnswnsn 225
QY 61 GTTGTAATACTATGATTAATGATCGATCGACACAGGATATTTATCATGTGGCACTCT 120
Qt 61 GTTGTAATACTATGATTAATGATCGATCGACACAGGATATTTATCATGTGGCACTCT 120
Db 226 mgnytngcngnngcaycctntaytawsgntnnaaraayayacnaa-racna- 283
Dt 226 mgnytngcngnngcaycctntaytawsgntnnaaraayayacnaa-racna- 283
QY 121 AGATTATTAACCTGTGGTATCAATATTTAGGTTCCTGCGAGTGGTGGCAATAGCAG 180
Qt 121 AGATTATTAACCTGTGGTATCAATATTTAGGTTCCTGCGAGTGGTGGCAATAGCAG 180
Db 284 -ayathcnaarngnswngcaycartaymngnttymngnngnmgntnccngayccn 342
Dt 284 -ayathcnaarngnswngcaycartaymngnttymngnngnmgntnccngayccn 342
QY 181 GATATTCCTTAAGTTTCTGCTACCAATATAGAGATATTTCCGGGTGAGTACCTGACCA 240
Qt 181 GATATTCCTTAAGTTTCTGCTACCAATATAGAGATATTTCCGGGTGAGTACCTGACCA 240
Db 1063 aaygggnathtygtgggnaaycarytnttygtngacngtngayacnncnngnwna 1122
Dt 1063 aaygggnathtygtgggnaaycarytnttygtngacngtngayacnncnngnwna 1122

```

```

Dt 343 aayaarttvggntnccngayacnaayathtaayaaccngaycargmngntngntgg 402
QY 241 AATAAATTTGGTTTACCTGATATAGTATTTAATCCTGAACACAACTTTAGTGTGG 300
Qt 241 AATAAATTTGGTTTACCTGATATAGTATTTAATCCTGAACACAACTTTAGTGTGG 300
Db 403 acvgglvgrgqpplgaglsgh 462
Dt 403 acvgglvgrgqpplgaglsgh 462
QY 301 GCGTGTGCGAGTGAATTTGCGGTGCTGACCTTTAGGTGTTGCGCTTGTGGCAT 360
Qt 301 GCGTGTGCGAGTGAATTTGCGGTGCTGACCTTTAGGTGTTGCGCTTGTGGCAT 360
Db 463 cnyntttaaayngntngayacngarwnswnaaytntngcnaayaayayvtnath 522
Dt 463 cnyntttaaayngntngayacngarwnswnaaytntngcnaayaayayvtnath 522
QY 361 CCATTTTAAATAAATAGTACACCTGAAGTTCCTCCATGCGCTACGCTAATGTTCT 420
Qt 361 CCATTTTAAATAAATAGTACACCTGAAGTTCCTCCATGCGCTACGCTAATGTTCT 420
Db 523 gargaywsmngayayaathwsngtngaygnaaacaracncarytntgyathngn 582
Dt 523 gargaywsmngayayaathwsngtngaygnaaacaracncarytntgyathngn 582
QY 421 GAGGACGTTAGGACAATGCTGCTGTAGATTATAAGCACACACAGTTATGTTTGGC 480
Qt 421 GAGGACGTTAGGACAATGCTGCTGTAGATTATAAGCACACACAGTTATGTTTGGC 480
Db 583 tgvacnccngcngatggngarcaraytgacnaaargngcngtntgyaarwnacncarg 642
Dt 583 tgvacnccngcngatggngarcaraytgacnaaargngcngtntgyaarwnacncarg 642
QY 481 TGTGCCCCCTATTGGGAACACTGGGCTAAAGCAGCTGCTGTAACCTGCTCCTTA 540
Qt 481 TGTGCCCCCTATTGGGAACACTGGGCTAAAGCAGCTGCTGTAACCTGCTCCTTA 540
Db 643 acnaccngngaytgyccnccnyntngnytnaayacnccnathargayngngaytg 702
Dt 643 acnaccngngaytgyccnccnyntngnytnaayacnccnathargayngngaytg 702
QY 541 TCACAGGGCGATGGCCCCCTTAGACITTAGACACACAGTTTGGAAAGTGGTATAG 600
Qt 541 TCACAGGGCGATGGCCCCCTTAGACITTAGACACACAGTTTGGAAAGTGGTATAG 600
Db 703 athgayaacngngcngatggngaytgyaarngnytnacargarwnacncargng 762
Dt 703 athgayaacngngcngatggngaytgyaarngnytnacargarwnacncargng 762
QY 601 GTAGATCTGGATGCTGCGGACCTTAGTACACTTCAGATTCAGATTAACCTAGGTA 660
Qt 601 GTAGATCTGGATGCTGCGGACCTTAGTACACTTCAGATTCAGATTAACCTAGGTA 660
Db 763 cnyntngaytngtncarwnacntgyaartaycngaytavytnaaratgswngay 822
Dt 763 cnyntngaytngtncarwnacntgyaartaycngaytavytnaaratgswngay 822
QY 661 CCATTCGATATGCTCAGCTATTGTAATATCCGATATTTACAAATGCTGCAGAT 720
Qt 661 CCATTCGATATGCTCAGCTATTGTAATATCCGATATTTACAAATGCTGCAGAT 720
Db 823 ayggdgs m w f y l r e q l f a h y 882
Dt 823 ayggdgs m w f y l r e q l f a h y 882
QY 721 CCTATGGGATTCATGCTTTTGGTACGCGGAGCAGCTTTTCTAGGCAATT 780
Qt 721 CCTATGGGATTCATGCTTTTGGTACGCGGAGCAGCTTTTCTAGGCAATT 780
Db 883 tyaaymngcngnaargtngngnngaracnathccngcngarytntaytnaargnsw 942
Dt 883 tyaaymngcngnaargtngngnngaracnathccngcngarytntaytnaargnsw 942
QY 781 TGAATAGGCGAGGACTATGGTGACACTGCTCCCTCACTTATATTAAGGCACA 840
Qt 781 TGAATAGGCGAGGACTATGGTGACACTGCTCCCTCACTTATATTAAGGCACA 840
Db 943 aayggmngarccnccnccnswngntngaytngacnccnccnswngnswngnath 1002
Dt 943 aayggmngarccnccnccnswngntngaytngacnccnccnswngnswngnath 1002
QY 841 GGTATCGGCTCACCTGCGAGCTGTGTATTCCTCCCTCCCAAGTGCCTATGTT 900
Qt 841 GGTATCGGCTCACCTGCGAGCTGTGTATTCCTCCCTCCCAAGTGCCTATGTT 900
Db 1003 acnswngcngcaycctnttyaayacncartaytgytncarngcncargncayay 1062
Dt 1003 acnswngcngcaycctnttyaayacncartaytgytncarngcncargncayay 1062
QY 901 ACCTTCGACTCCAGTTGTTTAAATACCATATTTGGTTACATTAAGCACAGGTCTAAC 960
Qt 901 ACCTTCGACTCCAGTTGTTTAAATACCATATTTGGTTACATTAAGCACAGGTCTAAC 960
Db 1063 aaygggnathtygtgggnaaycarytnttygtngacngtngayacnncnngnwna 1122
Dt 1063 aaygggnathtygtgggnaaycarytnttygtngacngtngayacnncnngnwna 1122

```













\*\*\*\*\*

RELEASE

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Mar 27 13:44:03 1997; MasPar time 1084.04 Seconds  
Tabular output not generated. 1157.881 Million cell updates/sec

Title: >US-08-409-122-1  
Description: (1-1524) from US08409122.seq  
Perfect Score: 1524  
N.A. Sequence: 1 ATGGCTTTGTGGCGGCTAG.....GTGTACGTGCGGAGGAGTAA 1524  
Comp: TACCGAAACCGCGGATC.....CACATGACGCTCTTCATT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 279077 seqs, 411808665 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-new5  
1.BCT 2.FUN 3:INV1 4:INV2 5:ORG 6:MAM 7:VRT 8:PLN 9:PRI  
10:PRO1 11:PRO2 12:ROD 13:SYN 14:UNC 15:VIR  
Database: genbank94

16:BCT1 17:BCT2 18:BCT3 19:BCT4 20:BCT5 21:BCT6 22:BCT7  
23:BCT8 24:INV1 25:INV2 26:INV3 28:INV4 29:INV5  
30:INV6 31:INV7 32:MAM1 33:MAM2 34:MAM3 35:VRT1 36:VRT2  
37:VRT3 38:PAT1 39:PAT2 40:PAT3 41:PHG 42:PLN1 43:PLN2  
44:PLN3 45:PLN4 46:PLN5 47:PLN6 48:PLN7 49:PLN8 50:PRI1  
51:PRI2 52:PRI3 53:PRI4 54:PRI5 55:PRI6 56:PRI7 57:PRI8  
58:PRI9 59:PRI10 60:PRI11 61:PRI12 62:PRI13 63:ROD1  
64:ROD2 65:ROD3 66:ROD4 67:ROD5 68:ROD6 69:ROD7 70:ROD8  
71:SYR 72:SYN 73:UNA 74:VRL1 75:VRL2 76:VRL3 77:VRL4  
78:VRL5 79:VRL6 80:VRL7 81:VRL8

Database: genbank-news

82:BCT 83:INV1 84:INV2 85:MAM 86:VRT 87:PAT 88:PHG

89:PLN 90:PRI 91:ROD 92:STR 93:SYN 94:UNA 95:VRL

Database: u-emb146\_94

96:part1

Statistics: Mean 11.592; Variance 4.289; scale 2.703

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	1484	97.4	7857	80	PAPHPV18 Human papillomavirus	0.00e+00
2	922	60.5	7858	79	HPV45 Human papillomavirus	0.00e+00
3	746	49.0	7896	79	HPV59VG Human papilloma virus	0.00e+00

	4	740	48.6	1855	79	HPV11	Human Papilloma Virus	0.00e+00
5	724	47.3	7905	79	HPU21941	Human papillomavirus	0.00e+00	
6	721	47.3	7833	80	PRT39	Human papillomavirus	0.00e+00	
7	718	47.1	6042	59	HOMHPME18	Human cellular DNA/Hu	0.00e+00	
8	535	35.1	7855	79	HPV26	Human papillomavirus	0.00e+00	
9	532	34.9	7916	79	HPU31784	Human papillomavirus	0.00e+00	
10	527	34.6	2920	15	HP37217	Human papillomavirus	0.00e+00	
11	527	34.6	2920	79	HPU37217	Human papillomavirus	0.00e+00	
12	527	34.6	7852	79	HPV30	Human papillomavirus	0.00e+00	
13	524	34.4	1517	40	113522	Sequence 1 from paten	0.00e+00	
14	525	34.4	7759	79	HPU37488	Human papillomavirus	0.00e+00	
15	515	33.8	1484	15	HP34168	Human papillomavirus	0.00e+00	
16	515	33.8	1484	79	HPU34185	Human papillomavirus	0.00e+00	
17	515	33.8	1484	15	HP34171	Human papillomavirus	0.00e+00	
18	515	33.8	1484	15	HP34181	Human papillomavirus	0.00e+00	
19	515	33.8	1484	79	HPU34181	Human papillomavirus	0.00e+00	
20	515	33.8	1484	15	HP34185	Human papillomavirus	0.00e+00	
21	515	33.8	1484	79	HPU34171	Human papillomavirus	0.00e+00	
22	515	33.8	1484	79	HPU34184	Human papillomavirus	0.00e+00	
23	515	33.8	1484	79	HPU34168	Human papillomavirus	0.00e+00	
24	515	33.8	1484	15	HP34169	Human papillomavirus	0.00e+00	
25	515	33.8	1484	79	HPU34169	Human papillomavirus	0.00e+00	
26	515	33.8	1484	15	HP34184	Human papillomavirus	0.00e+00	
27	514	33.7	1484	79	HPU34187	Human papillomavirus	0.00e+00	
28	514	33.7	1484	79	HPU34190	Human papillomavirus	0.00e+00	
29	514	33.7	1484	79	HPU34183	Human papillomavirus	0.00e+00	
30	514	33.7	1484	79	HPU34177	Human papillomavirus	0.00e+00	
31	514	33.7	1484	79	HPU34186	Human papillomavirus	0.00e+00	
32	514	33.7	1484	15	HP34186	Human papillomavirus	0.00e+00	
33	514	33.7	1484	79	HPU34166	Human papillomavirus	0.00e+00	
34	514	33.7	1484	79	HPU34175	Human papillomavirus	0.00e+00	
35	514	33.7	1484	15	HP34165	Human papillomavirus	0.00e+00	
36	514	33.7	1484	15	HP34177	Human papillomavirus	0.00e+00	
37	514	33.7	1484	15	HP34187	Human papillomavirus	0.00e+00	
38	514	33.7	1484	15	HP34175	Human papillomavirus	0.00e+00	
39	514	33.7	1484	15	HP34183	Human papillomavirus	0.00e+00	
40	514	33.7	1484	15	HP34190	Human papillomavirus	0.00e+00	
41	514	33.7	1484	79	HPU34165	Human papillomavirus	0.00e+00	
42	514	33.7	1484	15	HP34166	Human papillomavirus	0.00e+00	
43	513	33.7	1484	79	HPU34193	Human papillomavirus	0.00e+00	
44	513	33.7	1484	79	HPU34179	Human papillomavirus	0.00e+00	
45	513	33.7	1484	79	HPU34192	Human papillomavirus	0.00e+00	

## ALIGNMENTS

RESULT	1	PAPHPV18	7857 bp	DNA	VRL	12-SEP-1993
LOCUS		Human papillomavirus type 18 genome.				
DEFINITION		X05015				
ACCESSION		60975				
NID		genome; overlapping genes; unidentified reading frame.				
KEYWORDS		Human papillomavirus.				
SOURCE		Viruses; dsDNA viruses, no RNA stage; Papovaviridae;				
ORGANISM		Papillomavirus.				
REFERENCE		1 (bases 1 to 7857)				
AUTHORS		Cole, S.T. and Danos, O.				
TITLE		Nucleotide sequence and comparative analysis of the human				
JOURNAL		papillomavirus type 18 genome. Phylogeny of papillomaviruses and				
MEDLINE		repeated structure of the E6 and E7 gene products				
COMMENT		J. Mol. Biol. 193 (4), 599-608 (1987)				
FEATURES		87283882				
		Data kindly reviewed (14-AUG-1987) by Danos O.				
		Location/Qualifiers				
		source				
		1..7857				
		/organism="Human papillomavirus"				
		87				
		/note="E6 ORF start site"				
		105..581				
		/note="E6 (AA 1-158)"				
		/codon_start=1				
		/db_xref="PID:g60976"				

/translation="MARFEDPTRRYPKLPDLCTELNLSLQDIEITCVYCKTVLEITEV  
FEAFKDLVYVRSIPHAACHKCIDFYSRIRLHYSVSDVYGTLEKLTNTGLYNLL  
IRCLRCQKPLNPAEKRLHNEKRFRHNIAGHYRGQCHSCCNRRARQLORRRETQV"  
509  
/note="E7 ORF start site"  
590..907  
/note="E7 (AA 1-105)"  
/codon\_start=1  
/db\_xref="PID:g60977"  
/translation="MHGPQTADIVLHLEPONEIPVDLLCHBQLSDSEENDEIDGV  
HQHLPARRAEPQRTHMLCNCCKCEARIELVSESSADDLRAFQOLFNTLSFVPCWCA  
SQH"  
908  
/note="E1 ORF start site"  
914..2887  
/note="E1 (AA 1-657)"  
/codon\_start=1  
/db\_xref="PID:g60978"  
/translation="MADEGIDEGTGCNGWFIYVOAIVDKKTDGVIDSDDENATDIG  
SDWYDIDTQGTFCBOALETAQALPHAQEVHNDQVLRHVKRFAGGSTENSLPGER  
LEVDTLSPLQELISLNSGOKKAKRRLFTISDSGIGGSEVETATQIVTTNGSHGNVC  
SGSTSAIDNGTGENSSVDGSDNSNIENVPQCTIAQLKOLLKYNKQGMALNVP  
KDTYGLSFIDLVRNEKSKDTTCDWTAIFGVNPTIAEGKTLIQPFIYAHQCLDC  
KMGVLLIALRYKSKSLYVAGLSLTLHVPETCMLIQPKLRSSVAALYXWRTGIS  
NISEVGDTPWIKRTIIQRIQDSDSFDLSEMVQWAFDNELDESDMAFAYALLADS  
NSNAAPLNSNQCAQKILKDCATMKKHYYRAQRQNMNSQWIRPCSKIDDEGDMRPV  
QFYAQOIEFTFLGALKSELKGTPKKNCVFCGPANTGSKSYFGMSFIHFIOGAVPI  
VNSTSHFLEPITDKVAMLDADATTCWTYFTYMRNALDGNPISIDRKHKLPIQLKC  
PPILLTNTHPAKDRNPYLESRITVFEFNAFPFDRKNGNPVYEINDKNKCKFFERTW  
SRLDLHEEEDADTEGNPFGTFKLRAQONHRPL"  
2796  
/note="E2 ORF start site"  
2817..3914  
/note="E2 (AA 1-365)"  
/codon\_start=1  
/db\_xref="PID:g60979"  
/translation="MOTPKETLSRLSCVQDKRIIDHYENDSKDISQIQYWOLIRWEN  
AIFPAAREHIGITLHQVVPAYINISKSKAKHAILOMALQAGLSRYKTEWTLQDCH  
ELWANTPEQCFKKGQVQVYFDGNKNCMTYVAMDSVYMTDAGTWKTKATCVSHR  
GLYVREGNTYIFBKSECEKYGNTWEVHFNNVIDCNSMCSSTSDTYSATQLV  
KQLQHPSPYSSTVSGTAKTYGOTSAATPGHCGLAERKOHGCVNPLLAGATPGNN  
KRLKSGNNTPIIHLKGDRLSKCLRYLRKSHDYRDISSTWHWTGAGNEKTKILT  
3405..3409  
/note="E4 ORF start site"  
3418..3684  
/note="E4 (AA 1-88)"  
/codon\_start=1  
/db\_xref="PID:g60980"  
/translation="WILCAVPTVTRYPLLSLNSYSTPPHRIAPCPWPAPQPTARRR  
LLHDLTVDSRSSIVDSLHFSVLHLQATFKDGNVSVYLRL"  
3915  
/note="E5 ORF start site"  
3936..4157  
/note="E5 (AA 1-73)"  
/codon\_start=1  
/db\_xref="PID:g60981"  
/translation="MLSLIFLFCFCVCMVYVCHVPLPSVCMCAWLVFVYIVVIT  
SPATATVVFCLLPMLLHILHAILSLQ"  
4166  
/note="L2 ORF start site"  
4244..5632  
/note="L2 (AA 1-462)"  
/codon\_start=1  
/db\_xref="PID:g60982"  
/translation="MVSRRARRKRASVIDLKXCKQSGTCPPDPVPKVEGTTLADKI  
LQWSSLIGLGLGIGTSGGRTGYPLGGRSNTVYDVGTRPPVPVIEPVGPDPDS  
IVTLIEDSSVYVSGAPRTFTGSGEDITSAGTTTTPAVLDTFSSVSISNTFNP  
AFSDPIIYVQGTGEVAPVFPVGTSGHYEEIPLQTFASSGTGEETISSTPLPTV  
RRVAGPRLYSRAIQVSVANPEFTRPSLLTYDNPAPFVDVDTLAFDRPSVDPDSF  
MDIILRHRLPALSRGTQVFRSGQRATMFTKSGTIGARVHFYHDISFAPSPIE  
LQPLVATENDLDFIYADDDPAVPVPSRSTSFAPFKYSPITSSASSYSNTVPLT  
SSWDVPVYTGPDITLPTSTYSWPIVSPAPASTQYIGIHGTHYILWPLFYFIPKKRRR

VPYFFADGFWAA"  
5418  
/note="L1 ORF start site"  
5430..7136  
/note="L1 (AA 1-568)"  
/codon\_start=1  
/db\_xref="PID:g60983"  
/translation="MCLVTRVLIHLHLLPLXGPLYHPRPLHLSILVVMVHIIICGH  
YIILFNRNVNVPPIFLQWALWRPDSNTVYLPSPSVARVNVNTDDYVTPTSIFYHAGSSR  
LLTVGVEYRVPAGGNKQDIPKVSAYQVRFVRLPDPNKFGLPDTSYNPTQSLV  
WACAGVEIRGPOPLAGVLSGHPFYNKLDPTESHAAKTSNVEDVRDNVSDYVYKOTOLC  
ILGCAPAEIHEWAKGTACKRPLSQGDCPPELKNVLEDDGMDVTCYGAMDPTLQD  
TKCEVPLDICQICRYPDYLOMSADPYGDSMFECLRRQLFARHFNRRACGMGTVPQ  
SLVTKGTGMPASPGSCVYSPSPSGSIVTSDSOLFKNPYHLHKAQGHNNGYCWHNQLFV  
TVWDTTSPNLTICASTQSPVGGQYDATERKPKYSRHYEYDLOIFOLCITITLADVM  
SVTHSNSSILDEDMNFGYPPPTPSLVYARYVQSVAITCQKDAAPARENQDPTKLEF  
WVNDLKEKSLSLDDQYPLGRFLVQAGLRKRPITIGPRKRSAPSATTSKPAKVRVRA  
RK"

BASE COUNT 2365 a 1497 c 1680 g 2315 t  
ORIGIN

Query Match 97.48; Score 1484; DB 80; Length 7857;  
Best Local Similarity 98.7%; Pred. No. 0.00e+00;  
Matches 1504; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 5613 atggccttggggcgccctagtgacaataccgtatatctccacotcctctctgtggcaaga 5672  
QY 1 ATGGCTTTGTGGCGCCCTAGTGACAAATACCGTATACCTTCCACCTCCTCTTCTGTGGCAAGA 60

Db 5673 gttgtaaataccgatgattatgtgactccccacaagcatattttatcatctgctggcagctct 5732  
QY 61 GTTGTAATACTGATGATTATGTGACTCGCACAAACATATTTATCATCTGCGACGCTCT 120

Db 5733 agattattaacctgttggtaataccatatttaggggttcctgcaggtggtggcaataaagcag 5792  
QY 121 AGATTATTAACTGTGTGTAATCCATATTTTAGGGTTCCTGCGAGTGGTGCAATAAGCAG 180

Db 5793 gatattcctaaagtttctgcataccaatagatgatttaggggtgcagttacctgaccca 5852  
QY 181 GATATTCTTAAGGTTCTTCGATACCAATAGAGTATTTTCGGGTGCAGTACTGACCCA 240

Db 5853 aataaatttgggtttaccctgatactgattattatactctgaacacaaacagtttagtggg 5912  
QY 241 AATAAATTTGGTTTACCTGTATAATAGTATTTATATCTCTGAAACACACAGTTTAGTGTGG 300

Db 5913 gctctgtcgtggagtggaaattggcgcgtggtcagcctttagtggttggccttagtgggcat 5972  
QY 301 GCCTGTGCTGGAGTGGAAATTTGGCCGTGGTGCAGCCTTTTAGTGTTGGCCTTAGTGGCAT 360

Db 5973 ccattttataataattagatgacactgaaagtcccatcgccgcacgctctaattgttct 6032  
QY 361 CCATTTTATAATAATTAGATGACACTGAAAGTTCCTCATGCCGTACGCTAATGTTCT 420

Db 6033 gaggaagcttagggacaattgtctgtgattataagcagacacacagttatgtttgggc 6092  
QY 421 GAGGACGTTAGGGACAAATGTCTGTAGATTATAGCAGACACAGTTATGTATTGTTGGC 480

Db 6093 tgtgccccctgctattgggggaacactggggttaagcagcactggttgaatacgcgctcttta 6152  
QY 481 TGTCCCCCTGCTATTGGGGAAACACTGGGCTAAAGGCATGCTTGTAAATCGGCTCCTTAA 540

Db 6153 tcacagggcgattggccccctttagaaacttaaaacacagtttgggaagatggtgatag 6212  
QY 541 TCACAGGCGCATGGCCCCCTTTAGAACTTAAACACACAGTTTGGGAAGATGTTGATG 600

Db 6213 gtagatactgggatatggtgccaatggacttttagtacattgcaagatactaaatgtgaggtta 6272  
QY 601 GTAGATTGATATGGTGGCCATGGACTTTTAGTACATTGCAAGATACTAAATGTAGGTA 660

Db 6273 ccattggatattgtcagctcctatttggtaaatcctcgattatttacaatgtctcgagat 6332  
QY 661 CCATTGGATATTGTGACGTATTGTTGAAATATCTGTGATATTATACAAATGTCTGCAGAT 720

VPYFFADGFWAA"  
5418  
/note="L1 ORF start site"  
5430..7136  
/note="L1 (AA 1-568)"  
/codon\_start=1  
/db\_xref="PID:g60983"  
/translation="MCLVTRVLIHLHLLPLXGPLYHPRPLHLSILVVMVHIIICGH  
YIILFNRNVNVPPIFLQWALWRPDSNTVYLPSPSVARVNVNTDDYVTPTSIFYHAGSSR  
LLTVGVEYRVPAGGNKQDIPKVSAYQVRFVRLPDPNKFGLPDTSYNPTQSLV  
WACAGVEIRGPOPLAGVLSGHPFYNKLDPTESHAAKTSNVEDVRDNVSDYVYKOTOLC  
ILGCAPAEIHEWAKGTACKRPLSQGDCPPELKNVLEDDGMDVTCYGAMDPTLQD  
TKCEVPLDICQICRYPDYLOMSADPYGDSMFECLRRQLFARHFNRRACGMGTVPQ  
SLVTKGTGMPASPGSCVYSPSPSGSIVTSDSOLFKNPYHLHKAQGHNNGYCWHNQLFV  
TVWDTTSPNLTICASTQSPVGGQYDATERKPKYSRHYEYDLOIFOLCITITLADVM  
SVTHSNSSILDEDMNFGYPPPTPSLVYARYVQSVAITCQKDAAPARENQDPTKLEF  
WVNDLKEKSLSLDDQYPLGRFLVQAGLRKRPITIGPRKRSAPSATTSKPAKVRVRA  
RK"

BASE COUNT 2365 a 1497 c 1680 g 2315 t  
ORIGIN

Query Match 97.48; Score 1484; DB 80; Length 7857;  
Best Local Similarity 98.7%; Pred. No. 0.00e+00;  
Matches 1504; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 5613 atggccttggggcgccctagtgacaataccgtatatctccacotcctctctgtggcaaga 5672  
QY 1 ATGGCTTTGTGGCGCCCTAGTGACAAATACCGTATACCTTCCACCTCCTCTTCTGTGGCAAGA 60

Db 5673 gttgtaaataccgatgattatgtgactccccacaagcatattttatcatctgctggcagctct 5732  
QY 61 GTTGTAATACTGATGATTATGTGACTCGCACAAACATATTTATCATCTGCGACGCTCT 120

Db 5733 agattattaacctgttggtaataccatatttaggggttcctgcaggtggtggcaataaagcag 5792  
QY 121 AGATTATTAACTGTGTGTAATCCATATTTTAGGGTTCCTGCGAGTGGTGCAATAAGCAG 180

Db 5793 gatattcctaaagtttctgcataccaatagatgatttaggggtgcagttacctgaccca 5852  
QY 181 GATATTCTTAAGGTTCTTCGATACCAATAGAGTATTTTCGGGTGCAGTACTGACCCA 240

Db 5853 aataaatttgggtttaccctgatactgattattatactctgaacacaaacagtttagtggg 5912  
QY 241 AATAAATTTGGTTTACCTGTATAATAGTATTTATATCTCTGAAACACACAGTTTAGTGTGG 300

Db 5913 gctctgtcgtggagtggaaattggcgcgtggtcagcctttagtggttggccttagtgggcat 5972  
QY 301 GCCTGTGCTGGAGTGGAAATTTGGCCGTGGTGCAGCCTTTTAGTGTTGGCCTTAGTGGCAT 360

Db 5973 ccattttataataattagatgacactgaaagtcccatcgccgcacgctctaattgttct 6032  
QY 361 CCATTTTATAATAATTAGATGACACTGAAAGTTCCTCATGCCGTACGCTAATGTTCT 420

Db 6033 gaggaagcttagggacaattgtctgtgattataagcagacacacagttatgtttgggc 6092  
QY 421 GAGGACGTTAGGGACAAATGTCTGTAGATTATAGCAGACACAGTTATGTATTGTTGGC 480

Db 6093 tgtgccccctgctattgggggaacactggggttaagcagcactggttgaatacgcgctcttta 6152  
QY 481 TGTCCCCCTGCTATTGGGGAAACACTGGGCTAAAGGCATGCTTGTAAATCGGCTCCTTAA 540

Db 6153 tcacagggcgattggccccctttagaaacttaaaacacagtttgggaagatggtgatag 6212  
QY 541 TCACAGGCGCATGGCCCCCTTTAGAACTTAAACACACAGTTTGGGAAGATGTTGATG 600

Db 6213 gtagatactgggatatggtgccaatggacttttagtacattgcaagatactaaatgtgaggtta 6272  
QY 601 GTAGATTGATATGGTGGCCATGGACTTTTAGTACATTGCAAGATACTAAATGTAGGTA 660

Db 6273 ccattggatattgtcagctcctatttggtaaatcctcgattatttacaatgtctcgagat 6332  
QY 661 CCATTGGATATTGTGACGTATTGTTGAAATATCTGTGATATTATACAAATGTCTGCAGAT 720

Db 6333 cctatgggggattccattgtttttttgttactgagcggtgagcagctttttgttagcattttt 6392.  
|||||  
QY 721 CCTATGGGATTCATGTTTGTGTTTACGACGTCAGCAGCTTTTGTAGCATTTT 780  
Db 6393 tggaaatagacaggtactactggtgacacactgtccctcaatccttatattataaagcacaa 6452  
|||||  
QY 781 TGAATAGGACAGGTACTATGGGTGACACTGTCCTCAATCCTATATATAAAGGCACA 840  
Db 6453 ggtagctcttcacactgacagctgtgtgtattctccctctccaaagtggtctattgtt 6512  
|||||  
QY 841 GGTATGCTGCTTACCTGCGACCTGTGTGTTATCTCCCTCTCCAGTGGCTATTGTT 900  
Db 6513 acctgactcccgattgtttaataaacattatgtgtacataagcagggctataac 6572  
|||||  
QY 901 ACCTGACTCCCGATTGTTTAATAAACATTATGTTTACATTAAGGCACAGGTCATAAC 960  
Db 6573 aatggtgttggcgaataaataattattgttactgtggttagataccactccagtaac 6632  
|||||  
QY 961 AATGGTATCTGCTGGCATAATCAATATTGTTTACTGTGTGATACACCTCTGTATACC 1020  
Db 6633 aatttaacaatatgtcttctacacagctctcctgtacctggcggaataatgatgctacaaa 6692  
|||||  
QY 1021 AATTAAACAATATGCTTCTACAGTCTCTGTACCTGGGCAATATGATGTACAAA 1080  
Db 6693 ttttaagcagtagacagacatgttgaggaaatgattgagctttatttttcagttgtgt 6752  
|||||  
QY 1081 TTTAAGCAGTATAGCAGACATGTTGAAGATAATGATTGCGATTTATTTTCAGTTATGT 1140  
Db 6753 actattacttactgacagatgtattgctcctatatcattcatgattgagcagattttta 6812  
|||||  
QY 1141 ACTATTACTTTAACTCGAGATGTTATGCTCTATATTATCATAGTATGATGACGATTTTA 1200  
Db 6813 gagtagtggaactttgtgttcccccccccaactactagttgttggtacatatcgt 6872  
|||||  
QY 1201 GAGATGGAACTTTGGTGTCTCCCCCCCCCACTACTAGTTGTTGGATACATATCGT 1260  
Db 6873 ttgtacaactctgttctattacctgtcaaaaggaagtgcgcacccgctgaaataaggat 6932  
|||||  
QY 1261 TTGTACAATCTGTTGCTATTACCTGTCAAAAGGATGCTGCACCACTGAAATAAGGAT 1320  
Db 6933 cctatgataagtaagtttggaaatgtgattttaaagaaagttttcttagactta 6992  
|||||  
QY 1321 CCTATGATAAGTTAAAGTTTGGATGTGGATTTAAAGAAAGTTTCTTTTGGACTTA 1380  
Db 6993 gatcaatcccccttgagcgaataattttgttccagctgagttgctgcgaagccacc 7052  
|||||  
QY 1381 GATCAATATCCCTTGGACGTAATTTTGGTTTCAGCTGGATGCTGCACGCCACC 1440  
Db 7053 ataggccctcgaaactgtctccatctgccaactacgtcttctaaacctgccaagcgt 7112  
|||||  
QY 1441 ATAGGCCCTCGTAACGTTCTGCTCATCTGCCACTACGCTCTCTAAACCTGCCAAGCGT 1500  
Db 7113 gtgcgtgtaagtcagggaagtaa 7136  
|||||  
QY 1501 GTCGGTGTAGTCCGCAAGGAAGTAA 1524  
|||||

RESULT 2  
LOCUS HPV45 7858 bp DNA VRL 01-FEB-1995  
DEFINITION Human papillomavirus type 45 genomic DNA.  
ACCESSION X74478 X74778 X77480 X77481  
NID g397022  
KEYWORDS E1 gene; E2 gene; E4 gene; E6 gene; E7 gene; early protein;  
L1 gene; L2 gene; late protein.  
SOURCE Human papillomavirus type 45.  
ORGANISM Human papillomavirus type 45  
viruses; dsDNA viruses, no RNA stage; Papovaviridae;  
Papillomavirus.  
REFERENCE 1 (bases 1 to 7858)  
AUTHORS Delius.H. and Hofmann.B.  
TITLE Primer-directed sequencing of human papillomavirus types  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 7858)

AUTHORS Delius.H.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1993) to the EMBL/GenBank/DBJ databases. H.  
Delius, Deutsches Krebsforschungszentrum, Abteilung AIV, Im  
Neuenheimer Feld 506, W-6900 Heidelberg, FRG  
REFERENCE 3 (bases 1 to 7858)  
AUTHORS Delius.H. and Hofmann.B.  
TITLE Primer-directed sequencing of human papillomavirus types  
JOURNAL Curt. Top. Microbiol. Immunol. 186, 13-31 (1994)  
MEDLINE 94265501  
FEATURES  
Location/Qualifiers  
1..7858  
/organism="Human papillomavirus type 45"  
/clone="insert in HindIII site of pGEM-1"  
102..578  
/gene="E6"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:g397023"  
/translation="MARFDDPKRPPYKLPDLCTELNLSLOVSIACVTKATLRTVEY  
YQAFKDCIVVRCIAVAAKHCIDFYSRIPRELYXNSVYGELEKITTWELYNLL  
IRCLRCQPLPAERRLHKRRRPHSIAGTGTGOCNCTCCDQARERLRRTQV"  
587..907  
/gene="E7"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:g397024"  
/translation="MHGPRELQEIHLHPONELDPVDLLCYEQLSESEENDADG  
VSHAQLPARRAEPQRHILCVCCCKDRIELTVESSEDLRTLOQLFLSTLSFYCPWC  
ATNQ"  
914..2845  
/gene="E1"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:g397025"  
/translation="MADPEGTDEGTGCGNFFVETIVEKKTGDISSDEDEATDTG  
SDMVDITDLSICEQAQEQTAQALHAQEVQNDQAVLHLLKRFAGSKENSPLEQG  
LSVVDLPRLQELSLNKGHKAKRRLFTISDGYGCEVEAAETQVTVNTNAENGGS  
VHSTOSSGSDSDNAENDPHCSITELKELQAANKAAMLAFFDIYGLSTDLVRN  
PKSDTCTDWMALFQVNPVVAEGFKLIRPATILYAHIOCLDKWGLLALLRYKC  
KRNUTVAAGLSTLHVPETCMLEPPKLSRVAALYRTGISNISEVSGTPEWQ  
GLTIHQGIDSDNFDLMQVAFNDLTDESMAFYAQLACDNCNAAFLKSCQA  
RYKDCAVMCRHYKRAQKMMQWIKYKSKIDEGGWRPIVQFLRYGVFFSFL  
RALKFLAGTKPKKICLLYPANTCKTSYGMSEIHFLOGAIIISFVNSNSHFWLEPLAD  
TKVMLDDATCTCTVYFDNYAENALDGNPISIDRKHKLQKPCILLTSNIDPAKD  
KNWPLESRTVTFPFAFPDFKNGNPNVTEINDKNWKCFEFTWLSRLDHEDEADT  
EGIPGTFKCVTQNTPL"  
2769..3875  
/gene="E2"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:g397026"  
/translation="MKMTPKESLSERLSALQDKILDHYENDSKDINSQISWQLIRL  
ENALFTAREHITLNHVPPPINISKAHKAIELOMALKGLAQSYNNDEMTLQD  
TCEELWNPQCFKKGKTVHYFDGNKNCMVVWDSIYIIFETGWDRTAACVS  
YGVYIKDGTITYVQKSECKEYKGNSTWVOYGNVINDCNSDSTSDTSATO  
IVRQLQASTSTPTASVGTPKPHIQTPATRPCCGLTEQHHGVNTHVHPLCSS  
TSNNRRKVCSTPIIHLKGDKNLKLRLRYADHYSEISSSTHWTGCKNKTG  
ILTVYNSEVORNTFLDVVTIPNSVQISVGNTI"  
4236..5627  
/gene="L2"  
/codon\_start=1  
/product="late protein"  
/db\_xref="PID:g397028"  
/translation="MVSHRAARRKASATLYRTCKSGTCCPPDVINKVEGTTLADKI  
LQWSSGLIFGLGIGTGGSGGTGYVPLAGRSNTVVDVGPTRPVVIEPVPTDPS  
IVTLVDSVWASGAPVFTGTSGFTTSSTTAVLDITPTVDVSISSSTSTNP  
AFSDPSIIEVPTQGEVSGNIFGTETSGHGYEIPLOTFASSGSETPISSTPLTV  
RRVGRPLYSRANQVRYSTSQFLHPSLSLTFDNPAYEPLDTLSFPTSNVPSDF  
MDIIRLHRPALSSRGTVRFKRGATMTFRSGQIGRVHFIHDPIASATEIEL  
QPLISATNDSLDLFDPPAPTSTIHKFTYPSILTMPSTAASSSTNVPL  
TSADVPIYTGPDIIPLSPMTWPTSTPNASTTYYIGIHGTQYVWLYWYVYFKKRR  
RIPYFADGFVAA"

CDS 5530..7149  
/gene="L1"  
/codon\_start=1  
/product="late protein"  
/db\_xref="PID:q397029"  
/translation="MAHNIYGHGIIIFLNKNVPEIFLOMALWSPSTVYLPPSRV  
ARVSTDDYVSRISFHHAGSSRLITVGNPVFVNGNKNQAVPKVSAXOYVPRV  
ALPDNKKGLPDSITNPFORLVWACVMEIGSGOPLGIGLSGHPFNKLDPESEN  
ANAVITQDVNDVNDYQTOICLGVPAIGSHAKGTLCKPAQLQPCDCPPLEK  
NVIIDGQDVGVGNDVSDYDITCEVPLDICQSCRYDPLQMSADPFGDSMEFC  
LREQDPFHHNAGVYGTNPIDILINGISANREIFQSCYISFSGSIIISDSQ  
LEKHYVEDLQIFQLCILILAEVMSIHSNSSLLENWNFVPPPTISLVDITRF  
VGSVAVRQADSTPPEKQDFWDLKRAFSSDLQPLGRKFLVAGURRNP  
TGIPRRPAASTSPASPSNPKRVRIKKK"  
BASE COUNT 2409 a 1462 c 1652 g 2335 t  
ORIGIN  
Query Match 60.5%; Score 922; DB 79; Length 7858;  
Best Local Similarity 82.3%; Pred. No. 0.00e+00;  
Matches 1231; Conservative 0; Mismatches 255; Indels 9; Gaps 4;  
Db 5608 atggcttggggcgctagtgacagtcaggtatattcttccacacaccccttctgtggccaga 5667  
QY 1 ATGGCTTTGTGGCGGCTAGTGACAATACCGGTATACCTTCCACCTCCTCTGTGGCAAGA 60  
Db 5668 gtgtccagcactgatattatgtcttcgcacagcatattttatcatcgagcgagttcc 5727  
QY 61 GTTGTAAATACTGATGATTATGTGACTCGCAGCAGCATATTTATCATGTGGCAGCTCT 120  
Db 5728 cgaattaaactgaggaatccattatttaggttgtaccctaaatgggtgcaggttaataaa 5787  
QY 121 AGATTATTAACGTGTGTAATCAATATTTAGGGTT---CCTGAGGGTGTGGCAATAAG 177  
Db 5788 caggctgtctcaaggtatcccatcatcagcataggggtgttttagagtagctttaccogat 5847  
QY 178 CAGGATATTCCTAAGGTTTCTGCATACCAATATAGAGTATTTTCGGGTGCAGTTACCTGAC 237  
Db 5848 cctaataaatttgattacctgattctactatataatacctcgaacacacactttgggt 5907  
QY 238 CCANATAAATTTGGTTTACCTGATTAAGTATTATTAATCCCTGNAACACACGTTTAGTG 297  
Db 5908 tgggcattgtcagttgataaattggtgcgcagccacttttaggttagcttagtcgaatgac 5967  
QY 298 TGGGCGCTGCTGGAGTGGNAATTGCGCGTGGTCAGCCCTTAGGTGTTGGCCTTAGTGGG 357  
Db 5968 catccatttataaattggatgatacagaagaagtgctcatgcagctacagctgttatt 6027  
QY 358 CATCCATTTATATTAATATAGTACACTGAAAGTTCCTATGCCGCTACGTCATAGTT 417  
Db 6028 acgcaggatgttagggataatgtgcagttgattataagaacaacacagctgtgtattta 6087  
QY 418 TCTGAGGACGTTAGGACAAATGTCTGATTTAAGCAGACACAGTTATGTATTGTG 477  
Db 6088 ggtgtgtacgtgctattggtgagcactgggccaaggggcacacactttgtaaacctgcacaa 6147  
QY 478 GGCTGTGCCCTGCTATTGGGAACACACTGGGCTAAAGGCACACTGCTTGTAAATGCCGTCT 537  
Db 6148 ttgcacacctggtgactgctcctcttggaaacttaaaaaacacactattttaggagatggtgat 6207  
QY 538 TTATCACAGGCGGATTGCCCCCTTTAGAACTTAAGAACACACAGTTTGGAAAGATGTGTAT 597  
Db 6208 atggtggatcaggttatgggccaatgatttttagtacattgcaggtacacaagtcgag 6267  
QY 598 ATGTTAGATAGTATGATGTCACACTTTAGTACATTGCAAGATACATAAATGTGAG 657  
Db 6268 gtctcattagacattgtcaatccatcgttaaatccagattatttgcgaatgtctgt 6327  
QY 658 GTACCAATTGGATATTGTGAGTCTATTGTGTAATATCTTGATTATTTACAAATGTCTGCA 717  
Db 6328 gatccctatgggattctatgttttttggctacgcgcggaacaactgtttgcgaagacat 6387  
QY 718 GATCCTTATGGGATTCATGTTTTTTTGGCTTACGAGGTGAGCAGCTTTTGTGAGGCAT 777

Db 6388 ttttgaataggcgagtgattggtgacacagtcacctacagcctatatattaaagcc 6447  
QY 778 TTTTGAATAGGCGAGGTACTATGGTGACACTGTGCTCAATCCTTATATATTAAAGGC 837  
Db 6448 actagcgtaaatagsgtgaacccctggaggtgtgtattccctctccagtggtgc 6507  
QY 838 AC-AG-G---TAAGCGTCTCACTGGCAGCTGTGTATTCTCCCTCCCAAGTGGC 891  
Db 6508 tctattatactctgtattctctcaattatttaataagccattatggtttacataaagsgccag 6567  
QY 892 TCTATTGTACCTCTGACTCCAGTGTATTAAACCATATGTTTACATTAAGGCACAG 951  
Db 6568 ggcataacaatggtattgttggcataaatcagttgttctactgtagtgtgacactacc 6627  
QY 952 GGTATATAAATGATGTCTGGCATATCAATATTGTGTACTGTGTAGATACCACCT 1011  
Db 6628 cgcagtactaatttaacattatgttcctctacacaaatcctgtgtcccaagtacatatgac 6687  
QY 1012 CGTAGTACCAATTTAAACAATATGTGCTTCTACACAGTCTCCTGTACCTGGCAATATGAT 1071  
Db 6688 cctactaagtttaagcagctatagtagacatgtggaggaatatgatttacagttttattttt 6747  
QY 1072 GCTACCAAAATTAAGCAGTAGTAGCAGACATGTTGAAGAATATGATTTCAGTATTATTTT 1131  
Db 6748 cagttgtgcattattactttaactgcagaggttatgtcatatatccatagtatgaatagt 6807  
QY 1132 CAGTTATGTACTATTACTTTAACTGCAGATGTTATGCTCTATATTCATAGTATGAATAGC 1191  
Db 6808 agtatattgaaaaatttggaattttgttccctccaccctactacacagtttgggtggat 6867  
QY 1192 AGTATTTTAGAGGATTGGAACITTTGGTGTTCCTCCCGCCCAACTACTAGTTTGGTGGAT 1251  
Db 6868 acetatcgtttgtgcaactcagttcgtgttacctgtcaaaaggatactacacacctccagaa 6927  
QY 1252 ACATATCGTTTTGTACAACTGTGTCTATTACCTGTCAAAAGGATGCTGCACAGCTGAA 1311  
Db 6928 aagcaggatccatatgataataaagtatttggactgtgacctaaagaaaaaattttcc 6987  
QY 1312 AATAAGGATCCCTATGATAGTTAAGTTTGGATGTGGATTAAAGGAAAAGTTTCT 1371  
Db 6988 tccgatttggataaatcccttgggtggaagtttttagtccaggtcgggttagctcgt 7047  
QY 1372 TTGGACTTAGATCAATATCCCTTGGAGGTAAATTTTGGTTGAGGCTGGATTGGCTCGC 1431  
Db 7048 aggcctacacataggacactgtaaggtcctgctctccacgtctactgcactata 7102  
QY 1432 AAGCCACATAGGCCCTCGTAAACGTTCTGCTCCATCTGCCACTACGCTCTCTA 1486  
RESULT 3  
LOCUS HPV59VG 7896 bp DNA VRL 10-OCT-1994  
DEFINITION Human papilloma virus type 59, complete viral genome.  
ACCESSION X77858  
NID 9557236  
KEYWORDS complete genome.  
SOURCE Human papillomavirus type 59.  
ORGANISM Human papillomavirus type 59.  
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;  
Papillomavirus.  
REFERENCE 1 (bases 1 to 7896)  
AUTHORS Rho,J., Roy-Burman,A., Kim,H., de Villiers,E.M., Matsukura,T. and Choe,J.  
TITLE Nucleotide sequence and phylogenetic classification of human papillomavirus type 59  
JOURNAL Virology 203 (1), 158-161 (1994)  
MEDLINE 94303229  
REFERENCE 2 (bases 1 to 7896)  
AUTHORS Choe,J.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-1994) to the EMBL/GenBank/DBJ databases. J. Choe, C/O Hajo Delius, DKFZ - Abt. ATV, Im Neuenheimer Feld 506, 69120 Heidelberg, FRG



Location/Qualifiers		
1..7896	/db_xref="PID:g557244"	
/organism="Human papillomavirus type 59"	/translation="MALWRSSDNKYLPPSPVAKVYSTDEVVTRTSIFYHAGSRLLITVGHPEYKVPKGGNGRQDVPKVSAYQYFRVKLPDPNKGFLPONTYVDNPSQRLVWACVGVGEIGRCQPLGVGLSGHPLYNKLLDDTENSHVASAVDTKDRONVSYDYKQQLCIGCPVAIGEHWTKGTACKPTTVVQGDPPLEINTPIEDGMDVDTGYAMDYKQLQDNKSEVLDDICQICKPYDLOMSADAYGDSMFELRREOVFAHFWRNSGTGDLQPESLYIKTQDTRANPSYLYSPSPSGSVVTSOLFNPYWLHKAQGLNNGICWHNQFLVWQDTRSNLSVCASSTTSINPVYPTSEKFAVREVEFDLQFIQLCKITLITVMSYIHNNITILEDNFEGVTPPTASLVYRFSAAVTCOKDTAPPVODPDYDKLFWPVDLKERFSADLDQPLGRKLLQALGRPKPTIGPRKRAAPAPTSTPSPKRVKRSKR	/codon_start=1
CDS		
55..537	/db_xref="PID:g557238"	
/gene="orf putative E6"	/translation="MHGPKATLCDIVLDLEPQNYEEVDLYCYEQLPDSDSEKDEPDGVNHPLLARRAEPQRHNIVCCKCNQLQVYETSQDSGLRALQLFMDTLFSVCP	
CDS		
542...865	/db_xref="PID:g557237"	
/translation="MAREDEPTQRYKLPDLSSTLPIHDIRINCVCKEQLQEREF	/db_xref="PID:g557239"	
FEFANDLFIVRDTCTPAACLKICISYARVRLRYRDSVYGTLEAETKPLHELLIRCYRCLAPLCPTDKLKHITTEKRRFHNIAGIYTGCGRCRTRARHLRQQRASETLV	/translation="MADSEGTGEGTCNGHFFVQAIYDKKIGDKISDDDENATDTGSDLVDFIDDTTCVQARETAQALFNQEAQDAREMHVLRKFGCSINSESEKAAAGKAKSPLOEISVNVNHPKVRRLITVDSGYSEVMELETQVTVTNGDSNGSV	
CDS		
872...2806	/db_xref="PID:g557240"	
/gene="orf putative E1"	/translation="MQTVMDTLSQLSVLODQILEHYENDSKDINEHINYKLVRMENVILFAARENNIHLNHQVPELVSKNKAEEALEQSNRTSTYMPCLKHLFAGVCHS	
/codon_start=1	/db_xref="PID:g557241"	
/translation="MMPKNNGLQTSGRKIIMARLLIYMTLCVAVPTVTSKYPLDLLSNYHTPPRPKPTWAPKRGTVRRLESQDSVDHTSLSLPACQMTVITVTSQSSVCIQAT	/db_xref="PID:g557242"	
/translation="MIFLVFVCCVVCVCLVCCNPVLLQSYVMCAWTWLLVFIYVITSSYECFLYILFFIPLLLLYAHALLSIQ"	/db_xref="PID:g557243"	
/codon_start=1	/translation="MVSHRAARKRASATDLYKTCQAGTCDPSDINKVGTGLADKI	
/db_xref="PID:g557243"	/translation="MIFLVFVCCVVCVCLVCCNPVLLQSYVMCAWTWLLVFIYVITSSYECFLYILFFIPLLLLYAHALLSIQ"	
CDS		
4231...5625	/db_xref="PID:g557243"	
/gene="orf putative L2"	/translation="MVSHRAARKRASATDLYKTCQAGTCDPSDINKVGTGLADKI	
/codon_start=1	/db_xref="PID:g557243"	
/translation="MVSHRAARKRASATDLYKTCQAGTCDPSDINKVGTGLADKI	/db_xref="PID:g557243"	
CDS		
3908...4129	/db_xref="PID:g557242"	
/translation="MIFLVFVCCVVCVCLVCCNPVLLQSYVMCAWTWLLVFIYVITSSYECFLYILFFIPLLLLYAHALLSIQ"	/db_xref="PID:g557243"	
CDS		
5606...7132	/db_xref="PID:g557244"	
/translation="MALWRSSDNKYLPPSPVAKVYSTDEVVTRTSIFYHAGSRLLITVGHPEYKVPKGGNGRQDVPKVSAYQYFRVKLPDPNKGFLPONTYVDNPSQRLVWACVGVGEIGRCQPLGVGLSGHPLYNKLLDDTENSHVASAVDTKDRONVSYDYKQQLCIGCPVAIGEHWTKGTACKPTTVVQGDPPLEINTPIEDGMDVDTGYAMDYKQLQDNKSEVLDDICQICKPYDLOMSADAYGDSMFELRREOVFAHFWRNSGTGDLQPESLYIKTQDTRANPSYLYSPSPSGSVVTSOLFNPYWLHKAQGLNNGICWHNQFLVWQDTRSNLSVCASSTTSINPVYPTSEKFAVREVEFDLQFIQLCKITLITVMSYIHNNITILEDNFEGVTPPTASLVYRFSAAVTCOKDTAPPVODPDYDKLFWPVDLKERFSADLDQPLGRKLLQALGRPKPTIGPRKRAAPAPTSTPSPKRVKRSKR	/codon_start=1	
CDS		
BASE COUNT	2473 a 1457 c 1594 g 2372 t	
ORIGIN		
Query Match	49.0%; Score 746; DB 79; Length 7896;	
Best Local Similarity	75.4%; Pred. No. 0.00e+00;	
Matches 1106; Conservative	0; Mismatches 360; Indels 0; Gaps 0;	
Db	5606 atggtctatggcgcttctagtgacaacaaggtgtatctaccctccaccttcggttagctaag 5665	
QY	1 ATGCTTTGTGGCGCGCTAGTGACAATACCGTATACCTTCACCTCCTCTCTGTGCGCAGA 60	
Db	5666 gttgtagcagctgtagtgcacccgtagtaccagtagtattttaccacgagcaggttcc 5725	
QY	61 GTGTAAATACTGATGATTATGTGACTCGCACAGCATATTTTATCATGCTGGCAGCTCT 120	
Db	5726 agactttaccagttgacatccatattttaaaagtacctaagaagtggtaattggttagacag 5785	
QY	121 AGATTATTAATCTGTGGTATCCATATTTTAGGGTTCTTCGACGTTGGCAATAAGCAG 180	
Db	5786 gatgttctaaagtggtctgcatacaatacagagctatttaggggttaagttaccatgaccc 5845	
QY	181 GATATTCCTTAGGTTTCTGCATACCAATATAGATATTTCGGGTGCAGTACCTTACCTGACCCA 240	
Db	5846 aataaattggcctccagatacacacagtagtatgcctaaactcaactcaacgcttggttcgg 5905	
QY	241 AATAAATTGGTTTACCTGATAATAGTATTATTAATCTGAACACACAGCTTAGTGTGG 300	
Db	5906 gcctgtagtggttgaaatcggcgcgccacacttttaggggttagcactcagtggttcac 5965	
QY	301 GCCTGTCTCGAGTGGAAATGGCGGTGGTGCAGCTTTAGGTGTGGCCTTAGTGGGCAT 360	
Db	5966 ccattataataaattggatgacactgaaacctctcatgtagcatctgctgttgtagacc 6025	
QY	361 CCATTTATATAAATTAGTAGACACTGAAGTTCCCATGCCGTACGCTCTAATGTTCT 420	
Db	6026 aaagatacacgtagtgcctgtagtgcctgattataaaacacacacgctgtagtatttggc 6085	
QY	421 GAGGACGTTAGGGACAATGTCTGTAGATTATAAGCAGACACAGTATGATTTTGGGC 480	
Db	6086 tgttagcctgcattgagaaacactggacaaagggcactgctgttagagcctcactcgtg 6145	
QY	481 TGTGCCCCCTGCTATTGGGGAACACTGGGCTAAAGGCACTGCTTGTAAATCGCGTCCITTA 540	
Db	6146 gttcagggcgattgctccctccactagataaataaataacacacacacacacacacacacacacac 6205	
QY	541 TCACAGGCGGATGCCCCCTTTAGAACTTAAAGAACACAGTTTGGAGATGGTGATATG 600	
Db	6206 gtagacacagtagatggggtcctgactgacttataattgttcaggagataac	

```
QY 781 TGGATAGGCGAGTACTATGGGTGACACTGTGCCTCAATCTTATATATAAGGCACA 840
Db 6446 gacatcagtgccaaaccagcgagttatttatatccctcccccagtggtctgtggtt 6505
QY 841 GGTATGCGTCTTACCTGGCAGCTGTGTATTCTCCCTCCTCAAGTGGCTACTATTGT 900
Db 6506 actctgattccaaatattttaaaacacatattgctgcacaaagctcagggtttaaac 6565
QY 901 ACCTCTGACTCCAGTGTGTTAATAAACCATATTGGTTACATAAGGCACAGGGTCATAAC 960
Db 6566 aatgggtatgtttgcacaaatcattgttttaaacagtttagatatactactcagcacc 6625
QY 961 AATGGTATCTGCTGCGCATTAATCAATTATTGTTACTGTGTAGTACCACTCGTAGTACC 1020
Db 6626 aatcttttgtgtgtgttactactactcttcttacttacttacttacttacttacttact 6685
QY 1021 AATTTAAACAAATATGCTCTTACACAGTCTCTCTGTACCTGGGCAATATGATCTACDMAA 1080
Db 6586 tttaagaatatgcagacatctggtgggaatttgaattgtattgagtttatttcaactgtgt 6745
QY 1081 TTTAAGCAGTATAGCAGACATGTGTGAAGAAATATGATTTGACAGTTTATTTTTCAGTTATGT 1140
Db 6746 aaataaacattacacagaggttaattgctacatactaatgaataacacactattttt 6805
QY 1141 ACATTACTTTTAACTGCGAGATGTTATGCTCTATATTCATATGATGATGATGATGATTTTA 1200
Db 6806 gaggtatggaatttgggtgttacacacactcctactgtagtttagttgacacataccgt 6865
QY 1201 GAGGATTTGAACCTTTGGTGTTCCTCCCTCCCGCCCACTACTAGTTTGTGTGATACATATCGT 1260
Db 6866 ttgttcaatctgctgtctgaacttctcaaaaggacacccgcagcttaaacagagac 6925
QY 1261 TTTGTACAATCTGTGCTATTAGCTGTCAAAAGGATGCTGCACACAGCTGAAAAATAAGGAT 1320
Db 6926 ccttatgacaaactaaagtttggcctgtagatcttaagaagaagtttcttcagatctt 6985
QY 1321 CCCTATGATAAGCTTAAAGTTTGGAAATGTTGGATTTTAAAGGAAAAAGTTTCTTTGGACTTA 1380
Db 6986 gatcagttcccttggagcgtcaatttttatgcaattagagtagagtagagtagagtagag 7045
QY 1381 GATCAATATCCCTTGGAGCGTAAATTTTGGTTAGGCTGGATGCTGCGAAGCCACACC 1440
Db 7046 ataggccacgcaacgtgcagcgc 7071
QY 1441 ATAGGCCCTCGTAACGTTCTGTCTCC 1466
```

```
4
RESULT
LOCUS      HPV11      1855 bp      DNA      VRL      16-FEB-1996
DEFINITION      Human Papilloma Virus L1 gene for major capsid protein.
ACCESSION      X67161
NID      g1197494
KEYWORDS      major capsid protein.
SOURCE      Human papillomavirus type 68.
ORGANISM      Human papillomavirus type 68.
              Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
              Papillomavirus.
REFERENCE      1 (bases 1 to 1855)
AUTHORS      Longuet,M.
TITLE      Direct Submission
JOURNAL      Submitted (03-JUL-1992) M. Longuet, Institut Pasteur, Laboratoire
              des Papillomavirus, 25 Rue du Dr Roux, 75024 Paris Cedex 15, FRANCE
REMARK      Revised by author 05-DEC-95
REFERENCE      2 (bases 1 to 1855)
AUTHORS      Longuet,M., Beaudenon,S. and Orth,G.
TITLE      Two Novel Genital Human Papillomavirus (HPV) Types, HPV68 and
              HPV70, Related to the Potentially Oncogenic HPV39
JOURNAL      J. Clin. Microbiol. 34, 738-744 (1996)
FEATURES      Location/Qualifiers
              1..1855
              /organism="Human papillomavirus type 68"
              /cell_line="intraepithelial neoplasia"
              194..1711
              CDS
```

```
BASE COUNT      501 a      356 c      398 g      600 t
ORIGIN
Query Match      48.6%; Score 740; DB 79; Length 1855;
Best Local Similarity 75.0%; Pred. No. 0.00e+00;
Matches 1111; Conservative 0; Mismatches 371; Indels 0; Gaps 0;
Db 194 atggcattgtggcgagctagcgacacaatggtgtattgttcctcccccctcagtgcggaag 253
QY 1 ATGGCTTTGTGGCGCCCTAGTAGACATACCGTATACCTCCACCTCCTCTCTGTGGCAAGA 60
Db 254 gttgtcaatacagatgattatgtgacacgacctggcattgtattactatgctggtacatct 313
QY 61 GTTGTAAATACATGATGATTATGTGACTCGCACAAAGCATATTTTATCATGCTGGCAGCTCT 120
Db 314 aggttattaaactgtgagccatccatattttaaggttccctatgctctgggggcccgaagcag 373
QY 121 AGATTATTAACTGTGTGGTAATCCCATATTTTAGGGTTCTCGCAGGTGGTGGCAATAGCAG 180
Db 374 ggcattctaaagtgctcgcataatcaatacagagtggttttaggttaaccttacctgacct 433
QY 181 GATATTCCTAAGGTCTTCGCATACCAATATAGAGTATTTCGGGTGCGAGTTACCTGACCCA 240
Db 434 aataaaattagtgctcctgagctcacattatataatccagatatcacagcagcagtggtatgg 493
QY 241 AATAAAATTTGGTTTACCCTGATAATAGTATTATTAATCTCTGAAACACAACTTTAGTGTGG 300
Db 494 gccgtgtgtgtgtgaaataggtagggggcagccattggcgcttgcccttagtgggcat 553
QY 301 GCCGTGCTGAGTGGAAATTTGGCGGTGTCAGCCTTTTAGGTGTTGGCCTTAGTGGGCAT 360
Db 554 ccactataataagctggtgatgatactgaaattccccgttttctcctaataaaaatcct 613
QY 361 CCATTTTATAAATATAGATGACACTGAAAGTCCCATGCCGCTACGCTCTAATGTTCT 420
Db 614 aaacacagtaggacaatgttgagtgagtgactgtaaacacacacagctgtgtattataggc 673
QY 421 GAGGACCTTAGGACAAATGTGCTGTAGATTATAGACAGACACAGATTATGATTATTGGGC 480
Db 674 tgttctctgctattggcgagcactggggccaaagtaaatcttgaagcctaccatgta 733
QY 481 TGTCCCTCTCTATTGGGAAACACTGGCTTAAAGGCACTGCTTGTAAATCGCTCCCTTTA 540
Db 734 caacaaagggagctgtcccccatttggaaattggtaaatctcctctattgaggtatggcgatg 793
QY 541 TCACAGGGCGATTGCCCCCTTTTAACTTTAAGAACACACAGTTTGGGAAGATGTTGATG 600
Db 794 attgatacaggtatgtgtcattgagctattgtgtacattacaagaacacaaagcgaggtta 853
QY 601 STAGATCTCGATATGTTGGCTTGGCTTTAGTACATTTAGTACATGCAAGATATAAATGTGAGGTA 660
Db 854 cctttggatatgtcaatctggttgcacaaatctcctgactatttgcacaaatgtctgcagat 913
QY 661 CCATTTGGATATTTGTCAGTCTATTGTAATAATATCTTGATTTATTTACAAATGCTCTGCAGAT 720
Db 914 gttgtgagagcagtagtattttttgttttgcgttagggaacaggttatttgcaggcatttt 973
QY 721 CCTTATGGGATTTCCATGTTTTTTTGTCTTACGACGTGACGACGCTTTTGTGCTAGGCAATTT 780
Db 974 tggaaatagggagcagtggttagggggaacactattccaccagatgtattataaggggcact 1033
```

```
/gene="L1"
/codon_start=1
/product="major capsid protein"
/db_xref="PID:g1197495"
/translation="MALNRASDNNVYLPDPNFSPSTLYNPOTQRWVAC
VGHFVFPVMSGRKQIPKVSAYVYRVFVLPDPNFSPSTLYNPOTQRWVAC
VGFIIGSGQPLVGLSGHPLNRLDDTENSFSSNPKPDRDNVAVDCTOLCIIIG
CVPAIGHMAKSGSKPTNVQGGCPPELVNTPIEDGMIDTGYGDMFGTLOETKS
IKGDIICQSVCKYVDYLOMSADVYDGSDFLEFLREOLFAHFHNRGNGMVGDTIPTMY
LKTDDIETPTSSVYVAPSPSGSVSSDQLFNPKYWLHKAOGHNNICWHNOLFVW
DTSTPNTLTSTTDDTSPAVYDTSNKFKEVRYVEVDLQRTFOLCTILTSDVMSYI
HTMPALTLDNMGVAPPSPASLVYTYLQSAATTCOKDAPAPVKPDYDGLNFNV
DLKEKPSSELDQPLGRKFLQAGVRRRTIGPRKETATATTSTSKRKRYSK"
```



IVOLVESSVWSGTPITFTGTSGFEIISSTAITPAVLDITPASGVSQISTSYNPN  
 ADAPSLVETQGEVSGNIFVTPTSGHEIYDQFVNSHGTSQTEPTSTPVPFG  
 SVRAGPYSRATYQHVNNVEFVPSFSTFQNFAPFGSTLTPEDTAPDGF  
 LILHRLPALTSRGTVAFSRMLKNTMTFRGTQIGVHPIDISNTATEDIEM  
 QLLASSTEDGLYDADADINAMLHTSTGTSGPSHGLFVPSISTSYNNT  
 IFTPSTWDIYPTGPDVITPASPNLPEVPTSIDTITVAITQGSNLYLPLLYELK  
 KPEPTSTVEGTGVAV.

CDS

5590..7104

```

j000...r10"
/gene="L1"
/note="Method: conceptual translation supplied by author.;
putative L1"
/codon_start=1
/db_xref="procl160346"

```

[illegible]

BASE COUNT	2425 a	1495 c	1693 q	2292 t	KEFSTIEDQFPLGRKLELLVGARRKRTIGFRKKRPASAKSSSSASAKHKKRVSK
------------	--------	--------	--------	--------	--

Query Match 47.5%; Score 724; DB 79; Length 7905;  
Best Local Similarity 75.6%; Pred. No. 0.00e+00;  
Matches 1112; Conservative 0; Mismatches 352; Indels 6; Gaps 4

Db 5590 atggcttttgagggtctagtgaacaacaggtgtatttgcaccccccttcgtgaggaaag 5649  
|||||  
|||  
Qy 1 ATGGCTTTGTGGGGGCGCTAGTGACAATACCGTATACCTTCACACTCCTCTGTGGCAAGA 60

Db 5650 gttgtcaatacagatgattatgtaacacgtacaggcatatatattatcttgaagctct 5709  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 61 GTTGTAAATACATGATGATATGGACTCCACAAGCATATTTTATCATGTCTGGCAGCTCT 120

**Db**    5710 cgcttattaacagtaggcataccttattttaagggtacotgtaaagtggccgcgaagcacg    5769  
         + | ||||| + | ||||| ||||| ||||| ||||| ||||| |||||  
**Oy**    121 AGATTATTAACGTGGTAAATCCATAATTTTAGGTCCTGCAGGTGGTGCAATATGACGAG    180

Db	5770	gaaatacctaaggtgtctgcataatcagtatagggtatttaggggtatccccctacctgatacct	5829
Qy	181	GATATTCTTAAGGTTTTCGCATACCAATATAGATATTTCGGGGTGCAGTTACCTGACCCA	240

Db 5830 aataagtttggccttcggaatccttccctttaaatcctgacacacaacgcctggtatgg 5889  
||||| ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 241 AATAAATTTGGTTTACCTGATAATAGTATTATTAATCCTGAAACACAACGTTTAGTGTTGG 300

Db	5890	gcctgtatagtg	tggaattg	gtagagcc	agccattg	ggcgctt	ggcgtagtgacat	5949
Qy	301	GCCTGTGCTGGACTGGAAAT	TGGCCGTG	GTGAGCTTT	AGGTGTT	TGGCCTTAGTGGGCAT	360	

[illegible]

Db	6010	caggacagtagggacacaatgtgtctgtggactataagcacaacacagttatgtattataggc	6069
Qy	421	GAGGACGTTAGGGACAATGTGTCTGTAGATTATAGCAGACACACAGTTATGTTATTTGGGC	480

Db	6070	tgtgttctgctatgggagagcaactgggctaagggaagcctgtaagtc-ca-ctc-a	6125
Qv	481	TGFGCCCTGCTATTGGGAACACTGGGCTAAAGCACTGTTGTAAATCGGTCCTTTA	540

Db	6126	--acaggcgatgttcaccattagaattagttataactgcaattgaggatggcgaatatg	6183
Ov	541	TCACAGGCGGATGGCCCCCTTTAGAACTTAAGACACACAGTTTTGAAGATGGGTGATATG	600

Accession	Sequence	Position
Db	6184 a tagatacaggctatggtgccattcgtaattcgtaaacccaaagttaggta	6243
QY	601 GTAGATACTGGATATGGTGGCCATGGCACTTAGTACATTGAAGATACCTAACTGTGAGGTA	660

Db	6244	ccactagatttgcgcaatccgctgtaaatatctctgatttatttgcagatgctgtcgat	6303
QY	661	ccattggattatttgcacagctatttggtaaatatcctgattatttacaatattctgcagat	720
Db	6304	gtatatgggacagtgatgtttttttgttgcgaaggaaacagttgtttgccaagccattt	6363
QY	721	ccttatggggatttccatgtttttttgtgttgcagctgagcagctttttgtatggcatttt	780
Db	6364	tggaaatagagtgcatggttgggcagacaataacctctcagaggttatataaaggacag	6423
QY	781	tggaaataggcaggttactatgggtgacacgtgcctcaatccttatataaaggacaca	840
Db	6424	gatacagtgagcgtccctggctacacagtatatcccttcccnaagtgctctatggctc	6483
QY	841	ggtatggctgcttcacctggcagctgtgtgtatttctcccttccaaagtgctctattgt	900
Db	6484	tctctgatcccaagttgttaaaagccctattggttgcataaggcccgagacacaat	6543
QY	901	accctgcacgccagtggttttaaaaaccaatattggttacataaggcagcaggctataac	960
Db	6544	aatggcatttgttggcataaacagttgtttattactgttgggtggacatacacctagtact	6603
QY	961	aatggtatctcgtggcataaataattttgttactgtgtgtagatcacctcgtagtacc	1020
Db	6604	aattttacattgctcgtgcacgaaacggccatacctgcttatagccctacaaag	6663
QY	1021	aattttaacaatatgtccttctacacagctcctgttacctggcgcaatgatgatccaaa	1080
Db	6664	tttaaggaaatactaggcagtgtaggaataagattacaaatttatcttcaattgtgt	6723
QY	1081	tttaagcagtatagcagacatgttgaaataatgatttgcagttatttttccagttatgt	1140
Db	6724	actatcacattaactgctgagttatggctacatacctactatgaatcctgcaatttg	6783
QY	1141	actattactttactgcagatttatgtccttatattcatagtagtaacgataatttta	1200
Db	6784	gcaattggaaatagaggttacccttccacctgcgaagctgtggcagcagtatagg	6843
QY	1201	gaggattggaaacttttgggttcccccgcgcacactactagtttgggtgaatacatatcgt	1260
Db	6844	tatttcaaatcagcagctatagcatgtcaaaaggatgctcctacacctgaaaaaaggat	6903
QY	1261	tttgtgacaaactgtctctattacgtgcaaaaggatgtgcaccagctgaaaaataaggat	1320
Db	6904	ccctatgacgatttaaaatttggaaattgtttaaaggaaaaagtttagtacagaacta	6963
QY	1321	ccctatgataagtttaagtttttggaaattgtgatttttaaaggaaaaagtttttgggaattta	1380
Db	6964	gatcagtttcccttggggcgcaaatttttactacaggtaggggctcgacagctcctact	7023
QY	1381	gatcaaatcccttggacgtaaaatttttggtttggctgagctggattgcgtgcgaagccacc	1440
Db	7024	ataggccctcgcaaacgccttcgctcagct	7053
QY	1441	atagccctctgtaaacggttctcgtccatct	1470

RESULT	6				
LOCUS	PPHT39	7833 bp	DNA	VRL	22-MAR-1994
DEFINITION	Human papillomavirus ORFs.				
ACCESSION	M62849 M38185				
NID	g333245				
KEYWORDS	intraepithelial neoplasia; invasive carcinoma; papillomavirus.				
SOURCE	Human papillomavirus (type 39) DNA.				
ORGANISM	Human Papillomavirus (type 39)				
REFERENCE	1 (bases 1 to 7833)				
AUTHORS	Volpers,C. and Strecek,R.E.				
TITLE	Genome organization and nucleotide sequence of human papillomavirus type 39				
JOURNAL	Virology 181, 419-423 (1991)				
MEDLINE	91135017				
FEATURES	Location/Qualifiers				



```
QY 121 AGATTATTAACTCTTGTAATCCATATTTTAGGGTTCTTCGACAGTGGTGGCAATAAGCAG 180
Db 5823 gacattccaaagtgctcgcatacataataggtatttcgcgtgacattgcccgcattc 5882
QY 181 GATATCTCTAAGGTTCTGCATACCATAATAGAGTATTTCGGGTGCATTAACCTGACCA 240
Db 5883 aataaattcagttattccagatgcattccttatataatccagaacaacaaccttttagtatg 5942
QY 241 AATAAATTTGGTTACTCGTAATAGTATTATTAATCCGTAACACACGTTTAGTGGG 300
Db 5943 gcttgtagggtagggtagggtagggtagggtagggtagggtagggtagggtagggtaggg 6002
QY 301 GCGTGTGCTGGAGTGGAAATGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
Db 6003 ccattataataagacagagtagatgataagaaactcac---catttccatacaaccacaa 6059
QY 361 CCATTTTATATAATATAGTAGACACTGAAGTTCCTCCATGCCGCTAGCTAATGTTCT 420
Db 6060 aaggacagtagggataatgtctgtgtggattataaacagacacagtggtgctattatagc 6119
QY 421 GAGGACGTTAGGGACAATGTCTGTAGATTATAAGCACACACAGTTATGTTTGGGC 480
Db 6120 tgtgttccctccattggggagacactggggtagggtagggtagggtagggtagggtaggg 6179
QY 481 TGTGCGCCCTGCTATTGGGAACACTGGGCTAAAGGCACCTGCTTGTAAATCGCGTCT 540
Db 6180 tctacgggggactgctcctcttggaactagtaaacacccctattggagtaggtgatag 6239
QY 541 TCACAGGGCGATTGCCCCCTTTAGAACTTAAGAACACACAGTTTGGAAAGATGGTGATG 600
Db 6240 attgatactagcatagagctatgacttgggtgctgctattgcaggaacacaaagttagg 6299
QY 601 GFAGATCTGGATATGGTCCCATGGACCTTAGTACATTGCAAGATACTAAATGTGAGGTA 660
Db 6300 cctttagatattgtcaactcattgtgaacttagtaaacctcctgattatttgcgaatgtc 6359
QY 661 CCATTTGGATATTTGTCAGCTATTGTTAAATATCCTGATTAATTACAAATGCTGCA 720
Db 6360 tgtgtagggacagtagtctctcttcttcttagtaggaacaactgtttgcaaacatttt 6419
QY 721 CCTATGCGGATTCATGTTTTTTTGTCTAGACGCTGAGCAGCTTTTGTCTAGGCATTT 780
Db 6420 tggaaatcgtggtagtggtagggtagggtagggtagggtagggtagggtagggtagggtag 6479
QY 781 TGAATAGGGCAGTACTATGGGTGACACTGCTGCTCATATCTTATATTATTAAGGCACA 840
Db 6480 gatatacgtgcaaaccccggtggtctgtatctactgcccctctccacagcgggtccatgg 6539
QY 841 GGTATGCGTGCTTCACCTGGCAGCTGTGTGATTCCTCTCTCCAAAGTGGCTCTATTGTT 900
Db 6540 acctctgattccccagttatttaataagccttattggtctacataaaggcccagggccaca 6599
QY 901 ACCTCTGACTCCCAAGTTGTTTAAATAACCATATTGGTTACATAAGGCACAGGGTCA 960
Db 6600 aatggtagattgtggcataatcaattatttcttactgtgtgtgacactaccctgtagtacc 6659
QY 961 AATGGTATCTCTGGCATAATCAATTAATTTGTTACTGTGTAGATACCACCTCGTAGT 1020
Db 6660 aactttacattatctactcttagagctcttccactacactctctacatatgatccctctta 6719
QY 1021 AATTTAAACAATATGTCTTACACAGCTCCTGTACCTGGGCAATATGATGCTACCA 1080
Db 6720 ttttaaggaataaccagcagctggagtagtatttacaatttctatttcaactgtgt 6779
QY 1081 TTTAAGCAGTATAGCACATGTGTAAGAAATATGATTTTGCAGTATTTTTCAGTTATGT 1140
Db 6780 actgtcacattaaacaactatgttcttatattcacaactatgaattcctcttatgt 6839
QY 1141 ACTATTACTTTAAGTCAGATGTATGTCCTATATTTCATAGTAAATAGCAGTATTTTA 1200
Db 6840 gacaattggaatttctgttagctctccaccactctccagtttggtagacactacaga 6899
QY 1201 GAGGATTGGAACTTTGGTGTTCCTCCCTCCCGCCAACTACTAGTTTGGTGGATACATATCGT 1260
Db 6900 tacctacagttcgcagccattacatgtcaaaagatgctccagcacctgaaaaaagat 6959
QY 1261 TTGTACAAATCTGTGCTATTACCTGTCAAAAGGATGCTGCACAGCTGAAAAAAGGAT 1320
Db 6960 ccattagcaggtctaaagtttgggaattgacttaagggaagtttagtttggaaactt 7019
QY 1321 CCTATGATAGATTAAAGTTTGGAAATGCTGATTAAGGAAAAAGTTTCTTTGGACTTA 1380
Db 7020 gatactcccttggggagagtaaatatttttgcagcccaagggtccgcagggcgcact 7079
QY 1381 GATCAATATCCCTTGGACGTAAATTTTGGTTGAGCTGGATTCGGTCGCAAGCCACC 1440
Db 7080 ataggtccccaagaagcgcctgctgcacccacttctcgtctctc 7123
QY 1441 ATAGGCCCTCTAAACCTTCTGCTCATCTGCCACTAGCTCTC 1484
RESULT 7 HUMHPVME18 6042 bp DNA PRI 11-OCT-1991
LOCUS Human cellular DNA/Human papillomavirus proviral DNA.
ACCESSION M73258
NID 9184383
KEYWORDS proviral gene.
SOURCE Human papillomavirus DNA; and Homo sapiens cervix DNA.
ORGANISM Human papillomavirus
REFERENCE 1 (bases 1 to 6042)
AUTHORS Reuter,S., Delius,H., Kahn,T., Hofmann,B., Zur Hausen,H. and Schwarz,E.
TITLE Characterization of a novel human papillomavirus DNA in the cervical carcinoma cell line ME180
JOURNAL J. Virol. 65, 5564-5568 (1991)
MEDLINE 91374616
FEATURES
source 1..6042 Location/Qualifiers
source 1..5993 /organism="Human papillomavirus"
source 221..1633 /organism="Human papillomavirus"
mat_peptide 221..1633 /sequenced_mol="DNA"
mat_peptide /standard_name="L2"
mat_peptide 1584..3131 /codon_start=1
mat_peptide /standard_name="L1"
mat_peptide /note="Open reading frame"
mat_peptide 3925..4419 /codon_start=1
mat_peptide /standard_name="E5"
mat_peptide 4337..4759 /codon_start=1
mat_peptide /standard_name="E7"
mat_peptide 4763..5116 /codon_start=1
mat_peptide /standard_name="E1a"
mat_peptide 4960..6039 /codon_start=1
mat_peptide /standard_name="E1b"
source 5994..6042 /codon_start=1
source /organism="Homo sapiens"
source /cell_line="ME180"
source /sequenced_mol="DNA"
source /tissue_type="cervix"
BASE COUNT 1810 a 1155 c 1279 g 1798 t
ORIGIN
Query Match 47.1%; Score 718; DB 59; Length 6042;
Best Local Similarity 74.1%; Pred. No. 0.00e+00;
Matches 1100; Conservative 0; Mismatches 382; Indels 0; Gaps 0;
Db 1617 atggcatttgccgctctagcagacaacatgggtgatttggctccccctcagtgggcgaag 1676
```

1 ATGCTTTGGGGGCTAGTACAAATACCGTATACCTCCACCTCCTCTGTGGCAAGA 60  
1677 GTTGTCAATACAGATTACGTAACACGACTGGGCAATTTACTACTAGTCTGTACATCT 1736  
61 GTTGTAAATACATGATGATGACCTGCGCAACAGATATTTATCATGCTGGCAGCTCT 120  
1737 aggtattaaactagtagccatccatattttaaggtccctatgtctgtggggccgcaagcag 1796  
121 AGATTATTAACCTGTGGTAATCCATATATTTAGGGTTCTCGAGGTGGTGGCAATAGCAG 180  
1797 gacattccctaaggtgtctgcatatcaatacacagggtgttttaggatttccctacactgaccc 1856  
181 GATATTCCTAAGGTTCTGATACCAATATAGATATTTCCGGGTGCACTACCTGACCCA 240  
1857 aataaaatttagtcttcctgagctacattataaacctgatacgcagcgattggatgg 1916  
241 AATAAATTTGGTTACCTGATATAGTATTTATTAATCCTGAACACACACCTTTAGTGTGG 300  
1917 gcctgtgtgtgtgtaaatagtagggggcagccattaggtgtgtggtccttagtgggcat 1976  
301 GCCTGTGCTGAGTGGAAATGGCCGTGGTCAGCCTTTAGGTGTGGCCTTAGTGGGCAT 360  
1977 ccattataaataggctagatgatactgaataattcccccgttttctcccaaaaaaacct 2036  
361 CCATTTTATAATAAATTAGATGACACTGAAAGTTCCCATGCCGTAGCTCTAATGTTTCT 420  
2037 aaggacagtagggacaatttccagtgactataaacaagcaactatattatttaggc 2096  
421 GAGACGCTTAGGCAACATGCTGTAGATATTAAGCAGACACAGATTATGATTTTGGCG 480  
2097 tgtgttctgccattgggggagcactggccaaaggttaaatcttgaagccctagcaatgtg 2156  
481 TGTGCCCTGCTATTGGGGAACACTGGGCTAAGGCACCTGCTGTAATCGCTCCTTTA 540  
2157 cagccggggagctgtccaccattggaattagtaaatcacactattccggatgagcgatg 2216  
541 TCACAGGGCGGATTCGCCCTTTAGAACTTAAGAACACAGCTTTTGGAGAGATGTTGATG 600  
2217 attgatacagagatgtgtctagcttagcttagctacattacaagaacaaagcgaggtg 2276  
601 GTAGATACCTGGATGATGGTCCATGGACCTTTAGTACATGCAAGATATAATGTGAGGTA 660  
2277 ccttagatatgtcaatcagctgtcaaatatcctgactatttacaatgtctgcagat 2336  
661 CCATTGGATATTTGTCAGTCTATTGTGAATATCCTGATTATACAAATGCTGCAGAT 720  
2337 gtatatggagacagtagtcttcttctgttctacgtagggaacagttattgttaggcatttt 2396  
721 CCTATGGGATTCCTATGTTTTTTTGTCTACGAGCTGAGCAGCTTTTGTAGGCATTTT 780  
2397 tgaataagaggggagctgttaggggacactatacctactgaattgtatataggcgcaat 2456  
781 TGAATAGGCGAGGTACTATGGGTGACACTGTGCCTCAATCCTTATATATAAGGCACA 840  
2457 gacatcgtgacagctcctagtagttagtatgctcccccctgcctagtggtctatgta 2516  
841 GGTATGGCTGCTTCACCTGGCAGCTGTGTATTCTCCCTCTCCAAGTGGCTCTATTGTT 900  
2517 tctcagactcccagttatttaacaagccctattgtgtgcacaagggcacaggacacaac 2576  
901 ACCTCTGACCTCCAGTGTGTTTATAAACCATATTGTTTACATAGGCACAGGCTCATPAC 960  
2577 aatgggtattgtggcataatcaatttcttactgtgtgtgtgataccactcagctacc 2636  
961 AATGGTATCTGCTGGCATAATCAATTTTGTACTGTGGTAGATACCACTCGTAGTACC 1020  
2637 aatttacttcttactactgaatcagctagctgataccaaatatttatgatccctaaataa 2696  
1021 AATTACAAATATGTCTTACACAGCTCTCCGTGACCTGGGCAATATGATGTACCAAA 1080  
2697 tttaaaggaatatattaggcagtggtgaggaataatgatttggcaatttatatttcaggtgtg 2756

1081 TTAAACAGATATAGCACACATGTTGAAGAAATATGATTTCGAGTTTATTTTCAGTTATGT 1140  
2757 actataacattgtccactgatgtaagtctctatatacatatactaatcctcgtattttg 2816  
1141 ACTATTACTTTACTCGAGATGTTATGCTCTATATTCATATAGTAGTAATAGCAGTATTTTA 1200  
2817 gatgattggaattttggtgtgtccctccaccatctgctagctgtttagatatacataccgc 2876  
1201 GAGGATTGGAACCTTGGTGTCCCCGCCCACTACTAGTTTGGTGGATACATATCGT 1360  
2877 tatctcaatcacagcaaatatcatgtcaaaaagcgcctcaccactactataaaaggat 2936  
1261 TTTGTACAATCTTGTCTATTACCTGTCAAAAGGATGCTGCACAGCTGAAATAAAGAT 1320  
2937 ccataatgctgttaacttttgaatgtaatttaaggaagaaagtttagtctgaactg 2996  
1321 CCTATGATAGTAAAGTTTGGATGTTGATGTTAAAGGAAAGTTTCTTTGGACTTA 1380  
2997 gaccagttctcttaggacgcaaatcttcttacaggcaggtgtccgcgacgacacccact 3056  
1381 GATCAATATCCCTTGGACGTAATTTTGGTTCAGGCTGGATTGCTCGCAAGCCCAACC 1440  
3057 ataggccccgttaaacgctgtccacagcaactactgcatct 3098  
1441 ATAGGCCCTCGTAACGTTTCTGCTCATCTGCCACTACGTCT 1482

RESULT 8  
LOCUS HPV26 7855 bp DNA VRL 26-MAY-1994  
DEFINITION Human papillomavirus type 26 genomic DNA.  
ACCESSION X74472  
NID 9396956  
KEYWORDS E1 gene; E2 gene; E4 gene; E6 gene; E7 gene; early protein;  
L1 gene; L2 gene; late protein.  
SOURCE Human papillomavirus type 26.  
ORGANISM Human papillomavirus type 26  
Viruses; dsDNA viruses, no RNA stage; Papovaviridae;  
Papillomavirus.  
REFERENCE 1 (bases 1 to 7855)  
AUTHORS Dellus H. and Hofmann, B.  
TITLE Primer-directed sequencing of human papillomavirus types  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 7855)  
AUTHORS Dellus H.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1993) to the EMBL/GenBank/DBJ databases. H.  
Dellus, Deutsches Krebsforschungszentrum, Abteilung ATV, Im  
Neuenheimer Feld 506, W 6900 Heidelberg, FRG  
REFERENCE 3 (bases 1 to 7855)  
AUTHORS Dellus H. and Hofmann, B.  
TITLE Primer-directed sequencing of human papillomavirus types  
JOURNAL Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)  
MEDLINE 94265501  
FEATURES  
Location/Qualifiers  
1..7855  
/organism="Human papillomavirus type 26"  
/clone="Insert in BamHI site of pUC19"  
97..549  
/gene="E6"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:G396957"  
/translation="MFEDPRPRPTLHELCSLNTLQNVQCVCKETLQWADVYN  
FAICDLRVVDRSPYACKRCVYKTEYRYRVCSTVYGATLEALTKSLCNLLIR  
CHRCNPLGPEEKRIYDEKRRFHEIAGWKGKLCINWRPRRQTEQV"  
536..870  
/gene="E7"  
/codon\_start=1  
/product="early protein"  
/db\_xref="PID:G396958"  
/translation="MHGNIINIEDVILDVLPQPEIDLCRYEQLDYEQFSSDEDETQN  
MRDQARQAGQEVCEYRIEACQCCMNSIVQLAYOSSRQNVRLVLEQMLMEDVSLVCHQCA  
AQ"



```
CDS
878..2794
/gene="E1"
/codon_start=1
/product="early protein"
/db_xref="PID:g336959"
/translation="MDCEGNEEGRCQWSEVAIEKHTGDTISDDETNSDSTGS
RLGFDSSSDVAEQEVQALFOAQKQANTKAVNLRKLLGSONSPLODITNHC
DOOSQONTHOVANSQAKBRANDVSPGSGVTEVELPVOVDKQVFEENGIPVHC
SOGSASVEIDVDVUNSVQIQELLCNSVAAALLSKFKVGVSEFELVPRKS
DKCCSDWCAFAVGSARESKLIQYCLHYHIOCLNCNGVYVLLVRFCAKN
RTIKNCMLNPFQGLIPEPKURSTAVAFYIKIGLSNISEYIGDTPEWYRQI
QLHSDRYDFDSKVMCFADDDTDSEIAFIAQADIDSNAAAFKLSNCAQYI
KQALMTRYIKRAQRKSCMSQWQIQCKIEEGSWKEIAKFLRHQVHYEILQVL
AVLDATYSCMDYIDKIPNMGASQFAMSFIRKMGSVISYNSNSHFWLPQEDSL
LYXHSVTVPPNTPFPDSDNGNPNVYALIDVNMKSEFSTWLSRLDEADKENGEL
PAFKVPVGENTKLL"
/2736..3863

CDS
/gene="E2"
/codon_start=1
/product="early protein"
/db_xref="PID:g336960"
/translation="MENLCORLNACQEKILDYELNKLTDQIDWKLRYECAIFY
KAREGMCQINHOVVPSTVVCQKQAWAIEIHIALQSLINTDYNEAMTMRDTSYEM
MTEPKRCKKEGTVTVVFDCKNEMTMDYIRWKYVYKTDIGKCKGTGDVDAKIYIT
QYAGKYVDFEAEKYEIGTVQWAVHVGQVICPEFVSTCSNQISSTAKTAPYVS
NATQTEYVPEVGTETEARPYEKKRRLSGDPTVTVTVTAAATQPGQSVDTNN
NLHSTGGHGPRTDSSDQVFIHLKGDTSKLRYRFRKHKHGLYCNVSTWHWTS
NDNQOQVITIFNSITQRNFLTIVTKIPOSITSLGIMSL"
/4189..5607

CDS
/gene="L2"
/codon_start=1
/product="late protein"
/db_xref="PID:g336962"
/translation="MVAVRAPRRKRASATDLVTKCAAGTCPPDVIPKIEGSTLADKI
LQWSGQPLGLGIGTSGSGRTGYIPLGGGRPSVVDIGTRPRPIIIEYVPGTEP
SIVTLVESSIOGAPILPFGSGNGFELTSSATPAVLIDITPSATGVHTVNTNIN
PLIEPIDIQONAEQHLFTTSTAGTSHSEYIEMVEFASNGLEPSISNPIPG
IQRSVAPKQAOQVQKVDNPNIGNPSTFVFNENPAYSPIDETLTYSASTVADP
DELDIALHRNALSRGTLVYSLKOKATKATSGKIGATVHYVYDHSIQSPAEH
FEELIOLHSHALSFOLYADDPOTVSIHTPRMSYSTLPLVPRVYASVRESINT
STANVPLSIRSEFLPVYSGSDIYVTSSTPWSPLPPPTNLPVAVHGDNYILNYP
IYLIHRRKMPYEFSGDFVAY"
/5588..7099

CDS
/gene="L1"
/codon_start=1
/product="late protein"
/db_xref="PID:g336963"
/translation="MALWRKSDSKVILPPTPVSRVNVNDEVVTRTGIIYVAGSSRLLT
LGPYSIKTCQKAEIPKVSATQYRVFVHLPLDPNKFGLPDPQLINPDIERLVWAV
GVEVGRQPLIGLGLSHPLNKLDDTENSLELATVNAUDTNDNVSVDNKKQIOLCIIG
TPPLGHEWIGIGTKNTQIQGDCPPELISIIEDGMDITDGFAMWDETALQATKSD
VPIDISQCTKYFDYLRMSADTYSNMFELRQELFARHFNKAGAVGDAIPTFLYI
KGAEGRPEPTSSISYATPSGSMVTSQAQFNKPYMQRAQGHNGICRGNOLFVTCV
DTRTNLITISLSAASSTPFKPSDKQFIRHSEYELQIFOLCKITLITLDMAYI
HLMNASILEDWNFGTLPPPTASLEDAIRFKNSATTCORNAPPVPKEDPFQKFWDV
DLKEKFSIDLOFPLGRKFMLOAQIQRRPKLGTKRPLSSSTSSSTKRKKRLTK"
/2477 a 1449 c 1583 g 2346 t

BASE COUNT 2477 a 1449 c 1583 g 2346 t
ORIGIN

Query Match 35.18; Score 535; DB 79; Length 7855;
Best Local Similarity 69.59; Pred. No. 0.00e+00;
Matches 1007; Conservative 0; Mismatches 436; Indels 6; Gaps 5;

Db 5588 atggctttgtgagcagaggtatattctctccaccctgtgtctcg 5647
Qy 1 ATGGCTTTGTGGCGGCTAGTACATACCGTATACCTCCACCTCTCTGTGGCAAGA 60
Db 5648 gttgtcaacacggatgaatgtaactgcacacgcatattattatcgaggagctct 5707
Qy 61 GTTGTAAATAGTATGATGATGACTCGACACAGCATATTTTATCATGCTGGCAGCTCT 120
Db 5708 cgttttaacattaggacatccattttttccatacctaaactggc--caaaaggcgg 5765
```

```
Qy 121 AGATTTAACTGTGTGTAATCCATATTTAGGGTCTCTGCAGGTGGTGGCAATAAGCAG 180
Db 5766 aa-attcctaaggtatctgcctatcagtcacaggtatttagagtcacacctatgacct 5824
Qy 181 GATATTCCTAAGGTTCTGCATACCAATATAGATATTTCGGGTGCAGTTACTTGACCCA 240
Db 5825 aataaatttgatgctgcctgataccagttatataataatgcacacagaacgcctggtg 5884
Qy 241 AATAATTTGGTTTACTGTAATATAGTATTATATCTCTGMAACACACAGTTTAGTIGG 300
Db 5885 gctggtggtggttgtaagttggttagagagacagccatttaggccttagtgagacat 5944
Qy 301 CCTGCTGCTGGAGTGGAAATGCGCTGCTGACGCTTATAGTGTGCTTATAGTGGCAT 360
Db 5945 ccttgtttaataagttggatgtataccgaaaaactctattggcttactgtaaaaatgcagac 6004
Qy 361 CCATTTTATAATAATTAGATGACACTGAAGATTCCCAATGCCGCTACGCTCTAATGTTCT 420
Db 6005 actgcacacagggacaatggtttcagttgataataaaacacacagttatgtattataggt 6064
Qy 421 GAGGACGTTAGGGACAATGTGTCTGTAGATTATAAGCAGACACAGTTATGTATTTGGGC 480
Db 6065 tgcacacccctctggagagacactgggtgattggcactatagtataaatacacacagaca 6124
Qy 481 TGTGCCCTGCTATTGGGGAAACACTGGGCTTAAAGCACACTGCTGTAAATCGGCTCCTTTA 540
Db 6125 caacgtggggttgccccctcctagaataaattccagcattatttgaggtatgagcgtatg 6184
Qy 541 TCACAGGGCGATTGCCCCCTTTAGAACTTAAGAACACAGTTTGGAAAGATGCTGATATG 600
Db 6185 attgatacagggcttggagctatggattttactgcttcacaggtctacacaaatcagatgtg 6244
Qy 601 GTAGATACTGGATATGGTGGCATGGACTTTAGTACATTGCAAGATACTAAATGTGAGGTA 660
Db 6245 cccattgattatgacaatccactgaaatacctctgattatctctaaatgtctcagat 6304
Qy 661 CCATTGGATATTGTGACGTCTATTGTTAAATATCTGATTATTACAAATGTCGCAGAT 720
Db 6305 acatatggaacacagcatgtttttttctcccggaacaaattatttccagacatttt 6364
Qy 721 CCTATGGGGATTCACATGTTTTTGTCTTACGACGCTGAGCAGCTTTTGTAGGCATTTT 780
Db 6365 tataataagcggggggtgtggggatgctataccacacacttggtatataaagtgct 6424
Qy 781 TGGNATAGGGCAGGTACTATGGGTGACACTGTGCCCTCACTTATATATAAAGGCACA 840
Db 6425 gaatcaggcaggagccccctacatactttctattttctgtacacacctagtggtctctatg 6484
Qy 841 GG-T-ATGGGTCTTCACTCGGACGCTG-TGTGTAATCTCCCTCTCCAGTGGCTCAAT 897
Db 6485 gttactcgtgatagcacaactatttaataagcactactggttaaacgctgcacagggctcat 6544
Qy 898 GTTACTCTGACTCCCATGTTGTTTAAACCATATTTGGTTACATAAAGGCACAGGCTCAT 957
Db 6545 aataatggttatctgttgggggcaatcaattgtttgttacctgtgtgtataccaccgagct 6604
Qy 958 AACATGGTATCTGCTGCATATCAATATTATTGTTACTGTGTGTAGATACCACTCGTAGT 1017
Db 6605 actaacctaccattagttacattatctgcagcatctgcacactccactccatttaaacatct 6664
Qy 1018 ACCAATTTAAACAATATGCTCTCTACACAGTCTCTGTACCTGGGCAATATGATGCTACC 1077
Db 6665 gattataaacattttataagacatgaggagaataatgaattacaatttatattcagttg 6724
Qy 1078 AAATTTAAGCAGTATAGCAGACATGTTGAAGAATATGATTGTCAGTTTATTTTTCAGTTA 1137
Db 6725 tgtataataacacttacacacagatgttatggcttacacacatttaataatgaatgctccata 6784
Qy 1138 TGTACTATTACTTTAACTGACAGATGTTATGCTCTATTTCATATAGTATGAATAGCAGTAT 1197
Db 6785 ttgagaggttggaatttgggactaacctaccctccactgctagtttggagatgctcat 6844
```





```
Db 5813 atgcttctgagcctctagctacacacgctgtgtacctgctccacccccagctctcaaaa 5872
QY 1 ATGGCTTTTGGCGCCCTATGACATACCCGTATACCTTCCACCTCTCTGTGTGGCAGA 60
Db 5873 gttatcagcagcagcactatgtgacacgacacaaatattattattatcagggagcttt 5932
QY 61 GTTGTAAATACGTAGATATGTAGCTGCACACAGCATATTTATCATCTGTGGCAGCTCT 120
Db 5933 cgcctgctcatctggtgctacatccacattatccaatttcccaaaatccctcgtgtaaaagta 5992
QY 121 AGATTAACTAGTGTGAATCCCATATTTTAGGGTCTCTCAGGGTGGTGGCAATAGCAG 180
Db 5993 gatgtgcctcaagggtgtcgtcatttcagtcaggggtttccggtgctgttgcctgacacct 6052
QY 181 GATATTCTCAAGGTTTCTGCTACCAATATAGATATTTTCCGGTGCAGTTACCTGACCCA 240
Db 6053 aataagtttgggttgcgcgagtcgccgcataataaccctcagggcagaaagcttctggtg 6112
QY 241 AATAAATTTGGTGTACCTGATATAGTATTTATAATCTGTAACACACACAGTTTGTGTGG 300
Db 6113 gcttgcactggtgtggaggttaggtcgagggcgaacctctcgtgtgctgggttgagtggaac 6172
QY 301 GCCTGTGCTGGAGTGGAAATTGGCCGTGCTAGCCTTAGTGTGTGGCCTTAGTGGGCAT 360
Db 6173 cctctgtatacaacactgaatgacacagacaaactctaatttgcacatcctgaaatgggt 6232
QY 361 CCATTTTATAAATTAGATGACACTGAAGTTCCTCATGCCCTACGTCATAATGTTCT 420
Db 6233 caggattccagggacacattgctgttgactataagcaacacacactgctgcatctgggc 6292
QY 421 GAGGAGTTTAGGCAATGTCTGTAGATATATAGCAGACACAGTTATGTATTTTGGGC 480
Db 6293 tctacgctccatggcagaaactgggtgaagggcactgtgtgcacgacactagttcc 6352
QY 481 TGTGCCCTCTATTTGGGGAACACTGGGCTAAGGCACCTGTTGTAATTCGGCTCCTTTA 540
Db 6353 gctgctggtgattgccccccccctggagtttaattgacacacacataattggagatggcgatg 6412
QY 541 TCACAGGGCGATGGCCCCCTTTAGAACCTTAAGAACACAGTTTGAAGATGGTGATG 600
Db 6413 gtggataccgggtaagcgtgcacttggacttgcctgcgaagttaaatgaactcgtatgtg 6472
QY 601 GTAGATACGTGGATATGGTGGCATGGGACTTTAGTACATTTGCCAAGATACATAATGTGAGGTA 660
Db 6473 cccctgtatattgcccagctcagtgtaaatcctcgtactacttaggcagtggctgctgac 6532
QY 661 CCATTGGATATTTGTAGCTATTTGTAATAATCTGATTATTTACAAATGTCTGCAGAT 720
Db 6533 cccctatggcgacagcagctgtttttttctcgtgtaggggaacaaactgtttccagggcacttc 6592
QY 721 CCTATGGGGATTCCATGTTTCTTCTTACGACGTGAGCAGCTTTTGTCTAGGCATTT 780
Db 6593 tttaactcgtcgtgtgtagtaggggcaaaaatccccagattccttctacttaagggttaac 6652
QY 781 TGGAAATAGGCGCAGTACTATGGGTGACACTGTGCCTCAATCTTATATATTAAGGCACA 840
Db 6653 aacggcgagaaactccctggcagtgccatatacatcagctcccacacactagtgtggtccatgta 6712
QY 841 GGTATGCGTCTTACCTGGCAGCTGTGTATTCTCCCTCTCCAAAGTGGCTCTATTGTT 900
Db 6713 acctcagggctcaaatatttaataagccttactgctacagcagcccgacacacacac 6772
QY 901 ACCTCTGACTCCCACTGTTTATAAACCATATTGGTTACATAAGGCACAGGGTCAATAC 960
Db 6773 aatggttatgtgtggccaatcaggtatttttaactgtgtggacacacacaggaacc 6832
QY 961 AATGGTATCTGTGGCATATCAATATTTGTTGTTCTGTGTAGTACCCTCGTAGTACC 1020
Db 6833 aatatgtcgtgtgtgtacacacagctctcaaacctgttgacacacttatgatctaccacaa 6892
QY 1021 AATTTAACAATATGTCTCTACACAGTCTCTCTGTACCTGGGCATATGATGCTACCANA 1080
```

---

```
Db 6893 attaaagaatatttgacacatggggaggaatgatatttgcagtttattttccagttgct 6952
QY 1081 TTTAAGCAGTATAGCAGACATGTGAAGATATGATTTGCAGTTTATTTTCAGTTATGT 1140
Db 6953 aaagtctacattgacacctgaaatattggttacccttcatactatgaacagtcacctactt 7012
QY 1141 ACTATTACTTTAACTCGACAGATGTTATGCTCTATTTCAATATGAATAGCAGTATTTTA 1200
Db 7013 gaagactgaatttggattgacattgccaccttcacactagcttgggaagacagctatagg 7072
QY 1201 GAGATTGGAACTTTGGTGTTCCTCCCGCCCACTACTAGTTTGGTGGATACATATGCT 1260
Db 7073 ttgttaacatcctctgcataacttgcataaaagatttggccctcacagaaaagcagat 7132
QY 1261 TTTGTACAATCTGTTGCTATTACCTGTCAAAAAGATGCTGCACCAGCTGAAAATAAGGAT 1320
Db 7133 ccgtatgcataaactaaattctcgtggtatgattttaaaggatagattacccttggaattg 7192
QY 1321 CCTATGATAAAGTTTAAAGTTTGGAAATGTGGATTTTAAAGGAAAAGTTTCTTTGGACTTA 1380
Db 7193 tcacagttccctcgtggcagctaaatttttattacag 7228
QY 1381 GATCAATATCCCTTGGAGCTAAATTTTGGTTTCAG 1416
```

RESULT 10  
ID HP37217 standard; DNA; VRL; 2920 BP.  
AC U37217;  
DT 01-JAN-1996 (Rel. 46, Created)  
DE 27-MAR-1996 (Rel. 47, Last updated, Version 2)  
DE Human papillomavirus type 16 variant L1 and L2 capsid protein  
DE genes, complete cds.  
KW  
OS Human papillomavirus type 16  
OC Viridae; ds-DNA nonenveloped viruses; Papovaviridae;  
OC Papillomavirus.  
RV [1]  
RP 1-2920  
RA Icenogle J.P., Clancy K.A., Lin S.Y.;  
RT Sequence variation in the capsid protein genes of human  
RT Papillomavirus type 16 and type 31;  
RL Virology 214:664-669(1995).  
RN [2]  
RP 1-2920  
RA Icenogle J.P., Clancy K.A., Lin S.Y.;  
RT Submitted (28-SEP-1995) to the EMBL/GenBank/DBJ databases.  
RL Joseph P. Icenogle, Viral and Rickettsial Diseases, Centers for  
RL Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA  
RL 30333, USA  
CC NCBI gi: 1143286  
FH Key Location/Qualifiers  
FT source 1..2920  
FT /organism="Human papillomavirus type 16"  
FT /strain="1194"  
FT /note="variant; has 44 nucleotide differences in the capsid  
FT gene region compared to the accepted reference sequence,  
FT Genbank Accession Number K02718"  
FT CDS 1..1422  
FT /note="minor capsid protein; NCBI gi: 1143287"  
FT /codon\_start=1  
FT /product="L2 capsid protein"  
FT /db\_xref="PID:gi1143287"  
FT /translation="MRKRSARTRKASATOLYKCKQAGTCPPDIIPKVGKTIADQ  
FT I LQYSGMVFFGGLIGTGSCTGGRTGYIPLTRPPPTATDTLAPVRPPLTVDPVGPSPD  
FT S IVSLVEETSIDAGATPVPSPPDVSGFSITSTDTTPAIDINNVTWTTTTHNPT  
FT F TDPVSLQPPPAETGGGHFLLSSSTISTHNYEIPMDTFIVSNPNVTSSPIPGSRP  
FT V







Db	6072	gaggatagtaggacacattctctgtgatccaaagacaaacccaattgtgtattattggg	6131
QY	421	GAGGACGTTAGGGACAAATGTCGTAGATTATTAAGCAGACACAGATTATGTATTTTGGGC	480
Db	6132	tgactctctataggaacacattggctaaaggcaactgcctgtcttctgcgcacct	6191
QY	481	TGTGCCCTCTATTTGGGAACACTGGGCTAAAGGCACGTCTGTAAATGCGCTCTTTA	540
Db	6192	gcacaaggaggtctctctctggaactgttaattctctatacagagatggcgcaatg	6251
QY	541	TCACAGGGCGATTGCCCCCTTTAGACCTTAAAGACACAGTTTTGGAAAGTGTATG	600
Db	6252	gttgatattgggtttgttgcaactgattttaaacaattacaggaatctaataatcagatg	6311
QY	601	GTAGATAGTGGATATGGTCCCATGAGCTTTAGTACATTCGAAGATACTAATGTAGGTA	660
Db	6312	cttttagatatattcaatccacactgttaatatctctgattatttaaataatgagtcagac	6371
QY	661	CCATTGGATATTGTGAGTCTATTGTGTAATATCTCTGATTATTACAAATGTCTGCAGAT	720
Db	6372	gactatggcattctatgtgttctatttgcgtagggaacagttatttgcctaggcactac	6431
QY	721	CGTTATGGGATTTCCATGTGTTTTTGTCTACGACGTGAGCAGCTTTTGTAGGCATTTT	780
Db	6432	tttaaatgggcaggtgctatttggtaacaattaccctagcacattatatataaaagtaca	6491
QY	781	TGGAATAGGGCAGGTACTATGGGTGACACTGTGCCTCAATCCTTATATATAAAGGCACA	840
Db	6492	aataacagggatccccgcgaagctcagtatatgtgtactcctcctagtggtgctatggta	6551
QY	841	GGTATGCGTGCTTCACTGGCAGCTGTGTATTCTCCCTCCTCCAAAGTGCTCTATTGTT	900
Db	6552	acctctagagctcagttgtttaataaaccctactggttgcaacgcgcacagggacacat	6611
QY	901	ACCCTGACTCCCAGTTGTTTAAACCATATTTGTTTACATAAGGCACAGGTCATAC	960
Db	6612	aatggcatttgggtgggaacaggtatttgccttctgttggcaccactaggaacaca	6671
QY	961	AATGGTATCTGCTGGCATATCAATATTGTTACTGTGTAGTACCACTCTGTAGTACC	1020
Db	6672	caactgactactctgaacacacacaaagct-tat-cca- --ca-tataattcaacgca	6725
QY	1021	AATTTACATATATGCTCTACACAGCTCTCCTGTACCTGGGCAATATGATGTACCAA	1080
Db	6726	attaaacagtatgtaagaactgagaggaatatgaattcacagtttgtgttcaactgtgt	6785
QY	1081	TTTAGCAGTATAGCAGACATGTTGAAGATATGATTTGCAGTTTTATTTTTCAGTTATGT	1140
Db	6786	aaaatttcaactgtctgcagaaactatggcctattttacatactactatgaaactccactt	6845
QY	1141	ACTATTACTTTAACTGCAGATGTTATGTCCTATATTCTAGTATGTAATAGCAGTATTTTA	1200
Db	6846	gagggtggaatatgtggattgtcacccccagctgccacacagcttagaggacaaatcaga	6905
QY	1201	GAGGATTGGAACTTTGGTTTCCCCGCCGCCAACACTACTAGTTTGGTGGATACATATCGT	1260
Db	6906	tatgttaaagcccttgcataacctgcataaaggatcagccctcctgcgtgaaaaagggac	6965
QY	1261	TTTGTACAACTGTGTCTATTACCTGTCAAAAGGATGCTGCACCAAGCTGAAAATAAGGAT	1320
Db	6966	ccactagctaaaataaagtgttggatgttaactctacagacagattttctgtgcactgt	7025
QY	1321	CCCTATGATAAGTTAAAGTTTGGAAATGTGGATTAAAGGAAAAGTTTCTTTTGGACTTA	1380
Db	7026	gaccaattccactgggcagaaagtttttaatgcaacttggggttgcctactaaactct	7085
QY	1381	GATCATATCCCCCTGGACGTAAATTTTGTGTCAGGCTGGATTGCGTCGCAAGGCCACC	1440
Db	7086	actactactaaaagcgtcggcccccagttctctctacctctacacatcagcacaag	7144
QY	1441	ATAGGCCCTCGTAAAGGTTCTGCTCCATCTGCCACTACGCTCTCTTAACCTGCCAAGCG	1499
RESULT	13		
LOCUS	113522	1517 bp	DNA
DEFINITION	Sequence 1 from patent US 5437951.	PAT	02-AUG-1995
ACCESSION	113522		
NID	996326		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified		
REFERENCE	1 (bases 1 to 1517)		
AUTHORS	Lovy, B.R., Schiller, J.T. and Kirmbaur, R.		
TITLE	Self-assembling recombinant papillomavirus capsid proteins		
JOURNAL	Patent: US 5437951-A 1 01-AUG-1995;		
FEATURES	Location/Qualifiers		
source	1..1517		
BASE COUNT	487 a 293 c 284 g 453 t		
ORIGIN			
Query Match	34.4%;	Score 524;	DB 40; Length 1517;
Best Local Similarity	68.9%;	Pred. No. 0.00e+00;	
Matches	982; Conservative	0; Mismatches 440;	Indels 3; Gaps 2;
Db	1	atgtctcttggctgcttagtgaggccactgtctacttgcctcctctgtcccgatctctaaag	60
QY	1	ATGGCTTTGTGGCGCTAGTGACAAATACCGTATACCTTCCACCTCCTTCTGTGGCAAGA	60
Db	61	gttgaagcacggatgaatatgttcacgcacaaacacatatattatcatcgcaggaacatcc	120
QY	61	GTGTAAATACTGATGATTTATGTGACTCGCACAAAGCATATTTTATCATGCTGGCAGCTCT	120
Db	121	agactactgtcagttggagacccctctatttctctataaaaaacctaacaatacaaaaaa	180
QY	121	AGATTATTAACCTGTGGTAATCCATATTTTAGGGTTCCTCGAGGTGGTGGCAATAAGCAG	180
Db	181	ttagttctaaagtacaggtatcacatacagaggtatttagaatacatattacctgcacct	240
QY	181	GATATTCCTTAGGTTCTGCATACCAATATAGACTATTTTCGGGTGCAGTTACCTTGACCCA	240
Db	241	ataaagttgttttctgcacccctctatttataatccagatacacacagcgctggtttgg	300
QY	241	AATAAATTTGGTTTACCTGTAATAATAGTATTTTAACTCTGAAACACACACAGTTAGTGTGG	300
Db	301	gcctgtgaggtgttgaggtaggtcgtggtcagccattaggtgtggtggtcattagtggccat	360
QY	301	GCTGTGCTGGAGTGGAAATGGCCGTGTGAGCTTTAGGTGTGGCCTTAGTGTGGGCAT	360
Db	361	ctttattataaataattgtagtcacagaaatgctagtgtcttatgcagcaaatgcaggt	420
QY	361	CCATTTTATAATAATTAGATGACACTGAAAGTTCCCATGCGCTACGCTCAATGTTTCT	420
Db	421	gtggataatagagaatgtatatctatggtattacaaacacacacacacacacacacacac	480
QY	421	GAGGACGTTAGGGACAAATGTGTCTGTAGATTATTAAGCAGACACAGCTTATGTTTGGGC	480
Db	481	tgcacaaccacctaagggaacacactggggcaaggtatccccatgtaccacatgttgcagta	540
QY	481	TGTGCCCTGCTATTGGGGAACACTGGGCTAAAGGCACACTGCTTGTAAATCGCGTCCCTTA	540
Db	541	aatccaggtgattgtccaccattagagttataaacaacacagttattcaggatgggtgatg	600
QY	541	TCACAGGGCGATTGCCCCCTTTTAGAACTTAAGAACACAGTTTTTGGAAAGATGGTGATG	600
Db	601	gttcatactgcttgtgtgctagtgccttactacattacaggtcaacaaagatgaagtt	660
QY	601	GTAGATACTGGATATGTTGCCATGGACITTAGTACATTGCAAGATACTAAATGTGAGGTA	660
Db	661	ccactggatattgtacatctcttgcacaaatccagattattataaaatggtgtcgaa	720
QY	661	CCATTTGGATATTGTCAGTCTATTGTGTAATATCTCGATTATTTTACAAATGTCCTGCAGAT	720









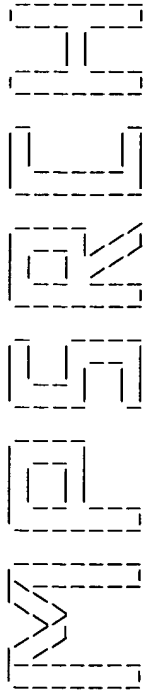
CC NCBI gi: 1098843 Location/Qualifiers  
 FH Key  
 FT source  
 FT 1..1484  
 FT /organism="Human papillomavirus"  
 FT /isolate="NM T446"  
 FT /note="collected by cervical swab at the University of New Mexico, New Mexico, United States."  
 FT CDS  
 FT <1..>1484  
 FT /gene="L1"  
 FT /codon\_start=1  
 FT /product="late major capsid protein L1"  
 FT /db\_xref="pid:gi098844"  
 FT /translation="TVLPPVPVSKVVDDEYVARTNIYYHAGTSRLLAYCHPFPPIK  
 K PNNKILPKVSGLOVRVRIHLDPNPKFGPDPSTFYNPDQRLWACVGEVGRGP  
 L GVIGSHPLNLKLDIDENASAYANAGVDNREISMDYKQTLCLGCKPPIGEHWGK  
 FT G SPTCNVAVNPGDCPPLELINTV IQDGDMDTGFAMDFTTLQANKSEVPDLICTSICK  
 FT Y PDYIKMVSPEYGDLSFPLRLREQMVFVRLFNAGAVGENVPDDLIIKSGSGSTANLASS  
 FT N YFTPPSGSMYTSDAQIFNKPYLQRAQHNGNIGWGNQLFVTVVDTRSTRNMSLCRAI  
 FT S TSETTYKNTNFKEYLRHGEYDLQIFQLCKITLITADVMTYIHSNMNSTILEDWNFGLQ  
 FT P PPGGTLEDYTRFVTSQAIAQCKHTPPAPKEDPLKKYTFWEVNLKXFSADLDQFPPLGR  
 FT K FLQAGLKAPKPTLGRKATPTSTSTAKRKRX"  
 FT S Q Sequence 1484 BP; 482 A; 285 C; 276 G; 441 T; 0 other;  
 Query Match 33.8%; Score 515; DB 15; Length 1484;  
 Best Local Similarity 69.0%; Pred. No. 0.00e+00;  
 Matches 964; Conservative 0; Mismatches 431; Indels 3; Gaps 2;  
 Db 1 actgtctactgctcctgtccagtagtcttaaggtgtgaagcagcgatgaatgtgtgca 60  
 Qy 28 ACCGTATACCTCCACCTCCCTCTGTGCAAGAGTGTGTAAATACTGATGATGTGACT 87  
 Db 61 cgcacaaacatattatcatcgaggaacatccagactactgactgagttggacatccctat 120  
 Qy 88 CGCAAGACATATTTATCATGCTGGCAGCTCTAGATTATTAACCTGTTGGTAATCCATAT 147  
 Db 121 ttctctataaaccacacaaataaataattagttctctaaagatcagggattacaa 180  
 Qy 148 TTTAGGTTCTCGAGGTGGTGGCAATAGCAGGATATTCCTAGGTTCTGCAATACCA 207  
 Db 181 tacaggggtattagaatacatttaccctgaccccaataagttttgttttccctgacacctca 240  
 Qy 208 TATAGATATTTCCGGTGCAGTTACCTGACCCAAATAAATTTGGTTTACCTGATAATAGT 267  
 Db 241 tttataatccagataacacagcgctggtttggccctgtgaggtgtgaggtgaggtcgt 300  
 Qy 268 ATTTATAATCTGAAACACACACGTTAGTGTGGCCCTGTGCTGGAGTGGAAATTTGGCCGT 327  
 Db 301 ggtcagccattaggtgtggtcattagtgccatctcttattataaataatttgatgaaca 360  
 Qy 328 GGTGAGCTTTAGGTGTGGCTTAGTGGGATCCATTTTATAATAATATAGATGACACT 387  
 Db 361 gaaaatgctagtctatcgacgaataatcaggtgtggataatagagaatgtatctatg 420  
 Qy 388 GAAAGTTCCCATGCGCTACGCTCAATGTTCTTGAGGACGTTAGGACAAATGTGCTCTA 447  
 Db 421 gattacaaacaaacaaattgtttaaattggttgcaaacaccatctataggggaacactgg 480  
 Qy 448 GATTATAAGCAGACACAGTATATTTTGGCTGTGCCCCGTGCTATTGGGGAAACACTGG 507  
 Db 481 ggcaagatcccatgtaccaaatgtgcagtaaaatccagtgattgtccaccattagag 540  
 Qy 508 GCTAAGGCACTGCTGTGTAATCGGCTCTTATCAGAGGCGGATGCCCCCTTTAGAA 567  
 Db 541 ttaataaacacagttattcaggatggtgatgattggtgactggtgcttgatgagac 600

Qy 568 CTTAAGAACACAGTTTGGAGATGGTCATATGTTAGTACTGATATGTTGCTCCATGAC 627  
 Db 601 ttactacatcagcgctaacaagaagttccactgataattgtacatctatttgc 660  
 Qy 628 TTTAGTACATGCAAGATACTAAATGTGAGTACCATTGGATATTGTGACTGTTTGT 687  
 Db 661 aaatccagattatattaaaatggtgtcagaaccatattgacgacagctattttttat 720  
 Qy 688 AATATCTGATTATTACAAATGCTGCAGATCCTTATGGGATTCATGTTTTTTCG 747  
 Db 721 ttacaaaggaacaaaatggtttgttagacattatttaataagggctggtgctggtgaa 780  
 Qy 748 TTACAGCTGAGCAGCTTTTGTGTCAGCAATTTTGGAAATAGGCGAGGTACTATGGTGAC 807  
 Db 781 aatgtacacagcatttatcattaaaagctctggtctactgcacaaatttagcagttca 840  
 Qy 808 ACTGTGCTCAATCTTATATATTAAAGGCACAGGTATGCGTCTTCACCTGGCAGCTGT 867  
 Db 841 aattatttctctacacctagtggttctatggttacctctgtagtgcacaaatattcaataa 900  
 Qy 868 GTGTATTCTCCCTCTCCAAGTGGCTCTATTGTPACCTCTGACTCCAGTGTGTTAATAA 927  
 Db 901 ccttattggtttacacagcagcagggccacaaataatggcattttgtgggttaacaaacta 960  
 Qy 928 CCATATTGGTTACATAAGGCACAGGTCATAACAATGGTATCTGCTGSCATAATCAATTA 987  
 Db 961 ttttactgtgtgtgatactacacgcagcagcacaataatgctcattatgctgctcattatct 1020  
 Qy 988 TTTGTTACTGTGGTAGATACCACTCGTAGTACCAATTTAAACATATATGCTTTCACACAG 1047  
 Db 1021 acttcagaaacta--ca--tataaaaataactaaactttaagggatcctcagcagatgggag 1077  
 Qy 1048 TCTCCTGTACTGGGCAATATGATGCTACCAAAATTTAAGCAGTATAGCAGACATGTTGAA 1107  
 Db 1078 gaatatgatttacagtttatttttcaactgtgcacaaataaacttaactgcagacggttatg 1137  
 Qy 1108 GAATGATGTTGCAAGTTTATTTTTCAGTTATGTACTATTACTTTAACTGCGAGATGTTATG 1167  
 Db 1138 acatacatattctatgaattccactattttggagactggaattttggtctcaacccc 1197  
 Qy 1168 TCCTATATTATAGTATGAATAGCAGATTTTATAGGATTGGAACTTTGGTGTCCCCCC 1227  
 Db 1198 ccccgaggagcagcactagaagatactttaggtttgttaacatcccaggaattgctgt 1257  
 Qy 1228 CCGCAACTACTACTGTTGGTGATACATATCGTTTTGTACAACTCTGTTGCTATTACCTGT 1287  
 Db 1258 caaaaaacatacacctccagcacctaaagaagatcccccttaaaaaaacacactttttggaa 1317  
 Qy 1288 CAAAAGGATGCTGCACGAGTGAATAAGGATCCCTATGATAAGTTAAAGTTTGGAAAT 1347  
 Db 1318 gtaaatttaaaggaagttttctgcagacctagatcagtttcccttaggacgcacaaattt 1377  
 Qy 1348 GTGGATTAAAGGAAAAGTTTCTTTGGACTTAGTCAATATCCCTTTGGACGTAAATTT 1407  
 Db 1378 ttactacaagcaggattg 1395  
 Qy 1408 TTGCTTCAGGCTGGATTG 1425

Search completed: Thu Mar 27 14:02:19 1997  
 Job time : 1096 secs.



\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, BioComputing Research Unit.

Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.

Distribution rights by IntelliGenetics, Inc.

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Mar 27 14:02:38 1997; MasPar time 154.35 Seconds  
861.234 Million cell updates/sec

Tabular output not generated.

Title: >US-08-409-122-1

Description: (1-1524) from US08409122.seq

Perfect Score: 1524

N.A. Sequence: 1 ATGGCTTTGTGGCGGCTAG.....GTGTACGTGCGAGGAAGTAA 1524

Comp: TACCGAAGACCGCGGATC.....CACATGACGCTTCCTCAT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 113505 seqs, 43611913 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-geneseq25

1:part2 2:part3 3:part4 4:part5 5:part6 6:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22

Statistics: Mean 9.390; Variance 5.312; scale 1.768

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	721	47.3	7833	4	Q25937 Human papilloma virus	0.00e-00
2	524	34.4	1517	7	BPV1 LI ORF.	0.00e-00
3	504	33.1	1517	7	HPV1 LI ORF.	0.00e-00
4	412	27.0	1599	20	T05834 Recombinant papilloma	8.17e-280
5	412	27.0	8010	18	T13413 Human papillomavirus	8.17e-280
6	403	26.4	7917	6	Q36030 Human papilloma virus	4.69e-273
7	380	24.9	2017	1	N91804 Human papilloma virus	8.50e-256
8	300	19.7	1950	1	Q04472 Human papilloma virus	5.51e-196
9	274	18.0	415	10	Q56380 HPV isolate C14 LI am	1.22e-176
10	274	18.0	415	1	Q03909 LI amplified region;	1.22e-176
11	266	17.5	416	17	T10811 Human papilloma virus	1.07e-170
12	251	16.5	455	17	T10805 Human papilloma virus	1.41e-159
13	251	16.5	455	10	Q56374 HPV isolate 36A LI am	1.41e-159
14	251	16.5	455	1	Q03903 LI amplified region;	1.41e-159
15	209	13.7	647	18	T03499 Papilloma virus major	1.42e-125
16	204	13.4	992	1	N91785 DNA probe complementa	6.72e-125
17	185	12.1	458	10	Q56377 HPV isolate 238A LI a	5.65e-111

18	185	12.1	458	17	T10808	Human papilloma virus	5.65e-111
19	185	12.1	458	1	Q03906	LI amplified region;	5.65e-111
20	179	11.7	3191	1	Q04092	3.2 kb KpnI-PvuII fra	1.37e-106
21	172	11.3	452	10	Q56378	HPV isolate 238B LI a	1.74e-101
22	172	11.3	452	1	Q03907	LI amplified region;	1.74e-101
23	172	11.3	452	17	T10809	Human papilloma virus	1.74e-101
24	157	10.3	450	10	Q56375	HPV isolate 36B LI am	1.37e-90
25	157	10.3	450	17	T10806	Human papilloma virus	1.37e-90
26	156	10.2	3637	11	Q03904	LI amplified region;	1.37e-90
27	156	10.2	3637	1	Q56372	BPV-1 polyfunctional	7.29e-90
28	154	10.1	2988	3	N50462	Sequence encoding bov	2.05e-88
29	145	9.5	450	17	T10807	Human papilloma virus	6.50e-82
30	145	9.5	450	10	Q56376	HPV isolate 88 LI amp	6.50e-82
31	141	9.3	450	1	Q03905	LI amplified region;	4.95e-79
32	140	9.2	2326	4	N30174	Sequence of the LI an	2.60e-78
33	130	8.5	412	10	Q56379	HPV isolate 155A/155B	3.90e-71
34	130	8.5	412	17	T10810	Human papilloma virus	3.90e-71
35	128	8.4	412	1	Q03908	LI amplified region;	1.05e-69
36	98	6.4	665	18	T03505	Papilloma virus major	1.78e-48
37	87	5.7	1047	2	Q10572	Human Natriuretic Pep	7.84e-41
38	83	5.4	661	18	T03501	Papilloma virus major	4.45e-38
39	78	5.1	1047	2	Q10572	Human Natriuretic Pep	1.17e-34
40	76	5.0	666	18	T03507	Papilloma virus major	2.69e-33
41	73	4.8	662	18	T03504	Papilloma virus major	2.90e-31
42	72	4.7	668	18	T03500	Papilloma virus major	1.37e-30
43	72	4.7	677	18	T03502	Papilloma virus major	1.37e-30
44	64	4.2	674	18	T03506	Papilloma virus major	3.08e-25
45	56	3.7	674	18	T03503	Papilloma virus major	5.42e-20

ALIGNMENTS

RESULT 1  
ID Q25937 standard; DNA; 7833 BP.  
AC Q25937;  
DT 18-JAN-1993 (first entry)  
DE Human papilloma virus HPV39 genome.  
KW HPV39 infection; cancer; glucocorticoid response element; GRE;  
KW enhancer; ss.  
OS Human papillomavirus 39.  
FH Key  
FT CDS Location/Qualifiers  
FT /\*tag= a complement (776..2050)  
FT CDS complement (3875..4204)  
FT /\*tag= b complement (406..411)  
FT /\*tag= c complement (424..427)  
FT TATA\_signal  
FT /\*tag= d complement (1861..1871)  
FT misc\_feature  
FT /\*tag= e  
FT /standard\_name= splice\_acceptor\_site  
FT /note= potential  
FT repeat\_unit 43..53  
FT /\*tag= f  
FT /function= enhancer  
FT /note= "papillomavirus-specific palindrome"  
FT repeat\_unit 58..70  
FT /\*tag= g  
FT /function= enhancer  
FT /note= "papillomavirus-specific palindrome"  
FT repeat\_unit 7456..7467  
FT /\*tag= h  
FT /function= enhancer  
FT /note= "papillomavirus-specific palindrome"  
FT WO9211369-A.  
PD 09-JUL-1992.  
PF 20-DEC-1991; F01053.  
PR 20-DEC-1990; PR-016044.  
PA (INSP) INST PASTEUR.  
PA (INRM) INST NAT SANTE & RECH MED.  
PI Orth G, Volpers C, Streeck RE;  
DR WPI; 92-250090/30.

PT DNA sequences of the HPV39 Papilloma: virus genome - used as a  
PT probe for in-vitro diagnosis of Papilloma: virus infections in  
PT cervical, vulvar and penile cancer

PS Claim 1; Fig 1; 3pp; French.

CC DNA was isolated from a previously identified HPV39 clone (Virology  
CC 161, 374-384, 1987), subcloned and sequenced. The HPV39 genome was  
CC found to comprise 7833bp and have a G/C content of 40%. The genome  
CC comprises open reading frames conserved among all the known HPV's;  
CC the ORFs are postulated to code for the early proteins and capsid  
CC components of the virus. In addition, two ORFs have been found on  
CC the complementary strand of the genome (see features table). The  
CC non-coding regions contain three complete copies of an HPV enhancer  
CC as well as two degenerate versions of the consensus palindromic  
CC (location not given). Other putative features which are described  
CC in the specification as being present within the regulatory region but  
CC whose positions are not given include: 4 nuclear factor-1 binding  
CC sites, 2 activation protein-1 binding sites, a papillomavirus  
CC enhancer associated factor binding site and a glucocorticoid  
CC response element. The HPV39 sequence has most homology with HPV18.  
CC The HPV39 is thought to belong to a sub-group of genital HPV's  
CC (with HPV18, HPV45 and a new type cloned from carcinoma ME180 cell  
CC line) representing a potential oncogene.  
SQ Sequence 7833 BP; 2423 A; 1486 C; 1662 G; 2262 T;

Query Match 47.3%; Score 721; DB 4; Length 7833;

Best Local Similarity 74.8%; Pred. No. 0.00e+00;

Matches 1110; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

Db 5643 atgggtatgtggcggtctagtagacagcagtggtgattgctccaccctctgtggcgag 5702

Qy 1 ATGGCTTTGTGGCGGCTAGTGACAAATACCGTATACCTCCACCTCCTCTGTGGCAAGA 60

Db 5703 gtgtgcataactgatgatgatgtacacgcacagcagcatatattatgtggcgagctct 5762

Qy 61 GTTGTAAATACGTATGATGATGTGCTCGCAAGCATATTTTATCATGCTGGCAGCTCT 120

Db 5763 agattattacacagcaggacacatcattttaagtggtgatgaatgggtggtgcgaagcag 5822

Qy 121 AGATTATTAACGTGTGTAATCCATATTTAGGCTTCCTGCAGGTGGTGGCAATAAGCAG 180

Db 5823 gacattcaaaaggtctcatatcaatagggtatttcggtgacattcccgatcct 5882

Qy 181 GATATTCCTAGGTTCTGCATACCANTATAGATATTTCGGTGCAGTTACTTGACCCA 240

Db 5883 aataaattcagattccagatgcattctatatataatccagaacacacacgttttagtagtg 5942

Qy 241 AATAAATTGGTTTACCTGATATAGTATTTATTAATCTGAAACACACGTTTATGTGG 300

Db 5943 gcttgtaggggtgaggtggtggcggggggccagccattgggtggttagtggtgacac 6002

Qy 301 GCCTGTGCTGGAGTGAATAATGGCCGTGGTGGCCTTTAGTGTGGCCTTAGTGGGCA 360

Db 6003 ccattataatagacagggatgatactgaaactcac---cattttcatcaaccacacaa 6059

Qy 361 CCATTTTATAAATATAGATGACACTGAAAGTTCCATGCGCTACGCTTAATGTTTCT 420

Db 6060 aaggacagtagggataatgctgtgattataaaccagacacagttgtgattataggc 6119

Qy 421 GAGGACGTTAGGGACAATGTGCTGTAGATTATAGCAGACACAGTTATGATTGTGGC 480

Db 6120 tgtgtcccgccattggggagcactgggtgaaaggaagcagtcgaagcccaataatgta 6179

Qy 481 TGTGCCCTGCTATTGGGAACACTGGGCTAAGGCACACTGCTGTGAATCGCTCTTA 540

Db 6180 tctacgggggactgtccctcttgaaactagtaaaacccctattgaggtggtgatg 6239

Qy 541 TCACAGGCGGATGCCCCCTTTAGAACTTAAGAACACAGTTTTTGGAAAGATGGTGATG 600

Db 6240 attgactgctgactgagctgacttggcttggctcaggaacacaaagtgaggtg 6299

Qy 601 GTAGATACTGGATATGTGGCAGTACCTTAGTACATTGCAAGATACATAATGTGAGGTA 660

Db 6300 ccttttagatattgtcaatccattgtgaaataatcctgattatttgcgaatgtgcagat 6359

Qy 661 CCATTGGATATTGTTCAGTCTATTTGTTAAATATCCTGATTATTACAAATGCTCAGAT 720

Db 6360 gtctatggggagcagtagtctctctgttttagtagtaggaacacacgttttgcagacatttt 6419

Qy 721 CCTATTGGGGATCCCATGTTTTTTTGTCTTACGACGCTGAGCAGCTTTTCTAGGCAATTT 780

Db 6420 tgaatactggtgtgtatggtggtgacgcacccattcctgcccaattgtatattaaaggca 6479

Qy 781 TGAATAGGGCNGGTACTATGGGTGACACTGTGCTCAATCCCTTATATTAAAGGCACA 840

Db 6480 gatatactggaaccccccggtagttctgtatctatctgccccctccccagcggttccatggta 6539

Qy 841 GGATACGGGTCTCACTGCGACGCTGTGTATTCTCCCTCTCCAAGTGGCTCTATTGTT 900

Db 6540 acctctgattcccgagttatttaataagcccttattggtcatataaagcccgagggccacaac 6599

Qy 901 ACCTCTGACCTCCAGTTGTTTAAATAACCATATTGGTTACATAAGGCACAGGGTCAATAAC 960

Db 6600 aatggtatattgtggcacaatacatatttcttactgtgtggtgacacactaccocgttagtacc 6659

Qy 961 AATGATATCTGCTGGCATAATCAATTATTGTTACTGTGTGTAGATACCACTCGTAGTACC 1020

Db 6660 aactttacattatctacacctatagagtgcttccatacctctctacatatgacctctctaa 6719

Qy 1021 AATTAAACAATATGTGCTTACACACAGTCTCCTGTACCTGGCAATATGATGCTACCAAA 1080

Db 6720 tttaaggaataaccagcgacgtggaggagtagtattttacataatttatttcaactgtgt 6779

Qy 1081 TTTAAGCAGTATAGCAGACATGTTGAAGAATATGATTGTCAGTTTATTTTTCAGTTATGT 1140

Db 6780 actgtcacattaaacaactgattgttcttcttatctcacactatgaattccctctattg 6839

Qy 1141 ACTATTACTTTAACTGACAGATGTTATGTCCTATTATTCATAGTATGAATAGCAGTATTTTA 1200

Db 6840 gacaaattggaattgtgtagctctccacacattccagtttctgtagacacttacaga 6899

Qy 1201 GAGGATTGGAACATTGTGTTCCTCCCGCCGCAACTACTAGTTGTGGATACATATCGT 1260

Db 6900 tactcacgtctgcagcattacatgtcaaaagactctccagcactccagcactcaaaagaagat 6959

Qy 1261 TTTGTACATCTGTGCTATTACTGTCAAAAGATGCTGCACCAGCTGAAAATAAGGAT 1320

Db 6960 ccattatgcggtctaaagtttggaaagtgtgacttaagggaagtttagtttggaactt 7019

Qy 1321 CCTATGATTAAGTTAAAGTTTGGAAITGGATTAAAGGAAGTTTCTTGGACTTA 1380

Db 7020 gatacaattcccttgggacgtataattttgttgagggccaggggtccgcagggcgccctact 7079

Qy 1381 GATCAATATCCCTTGGACGTAAATTTTGGTTAGGCTGGATTGGCTCGCAAGCCACC 1440

Db 7080 atagggtcccccgaagcgccctgctgcatccacttccctgctctc 7123

Qy 1441 ATAGGCCCTCGTAAACGTTCTGCTCCATCTGCCACTAGCTGCTTC 1484

RESULT 2

ID Q47166 standard; DNA; 1517 BP.

AC Q47166;

DT 21-JAN-1994 (first entry)

DE BPV1 L1 ORF.

KW L1; capsid protein; bovine; human; papillomavirus; baculovirus;

KW transfer vector; promoter; capsid protein; transformation; vaccine;

KW neutralising antibody; vertebrate; ss.

OS Bovine papillomavirus.

PN US8032859-A.

PD 15-JUL-1993.

PF 03-SEP-1992; 941371.

PR 03-SEP-1992; US-941371.

PR 16-MAR-1993; US-032869.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.

PI Kirmbauer R, Lowy DR, Schiller JT;

DR WPI; 93-249995/31.



Matches	972; Conservative	0; Mismatches	450; Indels	3; Gaps	2;	
Db	1	atgtctcttggcgtccagtagtgagggccactgtc	baettctgctccctgtcccaagtataag	60		
Qy	1	ATGGCTTTGTGGCGGCTAGTGACAAATACCGTATAC	CTTCCACCTCCCTCTGTGGGCAAGA	60		
Db	61	gttgaacacggatgaatatgttcgacgcacaaacata	tattatcatgcaggaacatcc	120		
Qy	61	GTGTGAAATACTGATGATGATGTACTCGCAACAG	TATTTTATCATGTCTGCGCAGCTCT	120		
Db	121	agactactgcagttgggacatccctattttctoctat	taaaaaacctaacaataa	180		
Qy	121	AGATTATTAACGTGTGTATTCATATTTTAGGTTCT	CGAGGTGGTGGCAATAAGCAG	180		
Db	181	ttagttctaaagtatcaggattacaacacaggggtat	ttagaatacatctacctgacccc	240		
Qy	181	GATATTTCTAAGGTTTCTGCATACCAATATAGAT	TATTTGGGTGTCAGTTACCTTGACCCA	240		
Db	241	ataagttgtgttctcgcacctcatcttataatccagat	tacacagcgcgtggtttg	300		
Qy	241	ATAAAATTTGGTTTACCTGATAATAGTATTATAAT	CCGTGAAACACACAGTTTAGTGTGG	300		
Db	301	gactgtagtggttaggttagtgctggtcagccattg	gtggtcagcttagtgagccat	360		
Qy	301	GCTTGCTGAGTGGAAATGGCGTGGTCAGCCITTAG	GTGTGGCTTAGTGGGCAT	360		
Db	361	ctttattataaataatggatgacacagaaatgctagt	gcttatagcagcaaatgcaggt	420		
Qy	361	CANITTTATAAATTAGATGACACTGAAAGTTTCC	CATGCGCTACGTCATTAATGTTCT	420		
Db	421	gtggataatagagaatgtatctatgttgatacaaa	caacaacatcttgtttaattg	480		
Qy	421	GAGGACGTTAGGGACAAATGTCTGTAGATTAT	TAAAGCACACAGATTATGTATTTGGGG	480		
Db	481	tgcaaacacacataggggaacactggggcaaggat	cccccatgccaatgttgagta	540		
Qy	481	TGTGCCCTGCTATTTGGGGACACATGGGCT	TAAAGGCACACTGCTGTGTAATCGCGCTTTA	540		
Db	541	aatccaggtgattgccaccattagtgtaataaac	acaggttattcaggatgggtatg	600		
Qy	541	TCACAGGGCGATTCGCCCTTTAGAACTTAA	GAACACAGTTTGGAAAGATGGTGATG	600		
Db	601	gttgatactggcttgggtgctagcttactacattac	agggctcaacaaagtggaagt	660		
Qy	601	GTAGATCTGGATATGTCGCATGGACTTTAGT	TACATTGCAAGATACATAATGTGAGGTA	660		
Db	661	caactggatattgtacatctatttgcacaaatcc	agattatataaaatggtgcagaa	720		
Qy	661	CCATTGGATATTTCAGTCTATTGTGTAATAT	CCTGATTATTTCAAATGTCGCGAGAT	720		
Db	721	caetatgcagcaccttatttttattatcagaagg	gaacaaatgttgtagacattta	780		
Qy	721	CCTATGGGGATCCAGTGTGTTTTTTCITAC	AGCTGAGCAGCTTTTGCTAGGCATTIT	780		
Db	781	tttaataaggctggtactgttgggaatgtaccag	agcagattatacataaaggctct	840		
Qy	781	TGGAATAGGGCAGGACTATGGGTGACCTGTG	CTCATCTTATATTAAGGCACA	840		
Db	841	gggtctactgcanaattbagcaggttcaaatatt	tttctcacaactagtggttctatggtt	900		
Qy	841	GGTATGGCGGTTCACCTGGCAGCTGTGTGAT	TCTCCCTCTCCAAGTGGGCTCATTTGTT	900		
Db	901	acctctgatgcccataattcaataaaccttattg	tttaacacagcagcagggccacaat	960		
Qy	901	ACCTCTGATCCCAAGTGTGTTTAATAAACCA	TATTTGGTTACATAAGGCACAGGGCTATAC	960		
Db	961	aatggcatttggggtaaccacatttgttactgtt	gttgatactacacgcagtcaca	1020		
Qy	961	AATGGGTATCTCGTGGCATAATCAATTAAT	TGTTTGTACTGTGGTAGATACCACTCGTAGTACC	1020		
Db	1021	aatatgcatatgtgtgcatactactctcaga	aacta--ca-tataaaaaa	lactaac	1077	
Qy	1021	AATTTAACAAATATGTGTTCTACACAGTCT	CCTCTACTCGGCAATATGATGCTACCAA		1080	

Db	153	agttgttcacagga	ctgtctatgttacttcgacacaaatattttatc	atgcacagcttc	212
QY	60	AGTTGAAAT	TACTGATGATTA	TGTGACTCGCAACAGCATATTTATCATGTCGCACTC	119
Db	213	tagactctcgag	tgggacatccctatttttccataaaa	-c-gg-gctaaca	aa-ct 267
QY	120	TAGATTA	TAACTGTGTGTA	TAACCATATTTTAGGGTCTCTGCAGGTGGTGCATAA	TAGCA 179
Db	268	gt-tgtgcaaa	aggtgtcaggatacaatcacagggtatttaaggtggtgttaccagatcc	326	
QY	180	GGATA	TTCTTAGGTTT	TCATACCAATATAGATA	TTTCGGGTGCAGTTACCTGACCC 239
Db	327	taacaaattg	caatgcctgaactctctcttcttcgaccccaacaaacaaagtttagtatg	386	
QY	240	AAATAAA	TTGGTTT	TACCTGATAA	TAGTATTTATAATCCTGAAACAAACGTTTAGTGTG 299
Db	387	ggcatcacag	ccctagaggtggcgagggacagccattagtggtgggtgaagtggaca	446	
QY	300	GGCCTGT	GTGAGTGGAAAT	TGGCGTGTGAGCCTTTAGTGTGTGGCCTTAGTGGGCA	359
Db	447	tcctttcc	taaatatgatgatgttgaaatcc--agggagtggtgg--taacctgtg	503	
QY	360	TCCATTT	TATAAAT	TAGATGACACTGAAAGTTC	CCACTGCCGTACGCTAATGTTTC 419
Db	504	acaggataca	agggttaactgtgtggtatggattataaaacaaacaaattatgcattgtgtg	563	
QY	420	TGAGGAC	GTTAGGACAATGTCTGTAGATTA	TATGACAGACAGATTA	TATGTTTGGG 479
Db	564	atgtgcccc	ctttggcgacattgggttaaggttaaacagtgtaactaacacactgt	623	
QY	480	CTGTGCC	CTGCTATTGGGGAACATGGCGCTTA	AGGCACTGCTGTGTAATCGCGCTCCTTT 539	
Db	624	acaggctgt	gactgcgcgccttagaacttattaccagtggttatcacaggatggcgat	683	
QY	540	ATCACAGG	CGGATGGCCCCCTTAGA	CTTAAGAACACAGCTTTGGNAGATGTGATAT 599	
Db	684	ggttacacag	ggttgggtgtatgaatttgcgtgatttcagagaccataaatacatgatgt	743	
QY	600	GGTAGAT	TACTGGATATG	TGTCATGGACTTTAGTAGATTTGCAAGTACTAAATCTGAGGT 659	
Db	744	tcctattgaca	tatggcactacatgtaataatccagattattitacaaatggcgtgcaga	803	
QY	660	ACCATT	TGGATATTGACATCTATTGTA	ATAATCCTGATTATTACAAATGCTCGAGA 719	
Db	804	cccatatg	gtgatagtattttttctacgggaagacaatgttttcgcagacattt	863	
QY	720	TCCTTAT	GGGATTCATGTTTTT	TGCTACGACGTGAGCAGCTTTTGTAGGCATTT 779	
Db	864	ttttaacag	ggtcggaggtggggaacctgtcctgatataactataaataaagggtag	923	
QY	780	TTGGAAT	TAGGCAGGTACTATGGT	GACACTGTGCTCAATCCTATATATTAAGGCAC 839	
Db	924	tggaaatc	gcagctctgtaggagtagtatataatgtaacccccggcgcctttcgtt	983	
QY	840	AGGTAT	TCGCTGCTACCTGGCAGCTGTGTATTTCTCCCTCTCCAA	TGGCTCTATGTT 899	
Db	984	gtcctctg	aggcacaattgttaataagcatttggctacaaaaagccagggacataa	1043	
QY	900	TACCTCT	GACTCCAGCTGTTTAA	TAAACCATATTTGGTTACATAGGCACAGGTCATAA 959	
Db	1044	caatggtatt	gttggggaatacaactgtttgtactgtggtagataccacacgcagtac	1103	
QY	960	CAAT	TGTATCTGGCATAATCAATTA	TTTGTACTGTGTAGATACCACTCGTAGTAC 1019	
Db	1104	caacatga	caattatgtgcatcogta-acta-catcttcca--ca-tacaccaattctga	1157	
QY	1020	CAATTT	TAAACAATATGTGCTCTAC	ACAGTCTCCTGTACTCTGGGCAATATGATGTACCAA 1079	
Db	1158	ttataaag	agtagcatcgtcatgtggaagatgatatttacaatttatttttcattatg	1217	
QY	1080	ATTTAG	CAAGTATGACAGACATGTTG	AGAAATATGATTTGCAGGTTATTTTTTCAGTTATG 1139	
Db	1218	tagcattac	attctgtcgtgaaqtaatggcctatttacacaaatgaatccctctgtttt	1277	

RESULT	5	
ID	T13413	standard; cDNA; 8010 BP.
AC	T13413;	
DT	12-JUN-1996	(first entry)
DE	Human papillomavirus type 6a.	
KW	vaccine; immunotherapy; genetic immunisation; condylo	ma acuminata;
KW	ss.	
OS	Human papillomavirus type 6a.	
EH	Key	Location/Qualifiers
FT	CDS	31..555
FT	FT	/*tag= a
FT	FT	/label= ORF-E6
FT	CDS	441..827
FT	FT	/*tag= b
FT	FT	/label= ORF-E7
FT	CDS	716..2782
FT	FT	/*tag= c
FT	FT	/label= ORF-E1
FT	CDS	2695..3830
FT	FT	/*tag= d
FT	FT	/label= ORF-E2
FT	CDS	2240..3585
FT	FT	/*tag= e
FT	FT	/label= ORF-E4
FT	CDS	3888..4163
FT	FT	/*tag= f
FT	FT	/label= ORF-E5
FT	CDS	4379..5803
FT	FT	/*tag= g
FT	FT	/label= ORF-L2
FT	CDS	5679..7292
FT	FT	/*tag= h
FT	FT	/label= ORF-L1
FT	FT	/product= major capsid protein
PN	WO9609375-A1.	
PD	28-MAR-1996.	
PR	18-SEP-1995;	U11859.
PR	22-SEP-1994;	US-310468.
PA	(MERI )	MERCK & CO INC.
PI	Hofmann KJ,	Jansen KU;
DR	WPI; 96-188437/19.	
PT	DNA encoding human papillomavirus 6a	- used to treat and in
PT	PT	vaccines against human papillomavirus 6a disease, e.g. condyloma
PT	PT	acuminata
PS	Claim 1; Fig 1A-G; 46pp;	English.
CC	A DNA molecule (T13413) codes for human papillomavirus 6a (HPV6a),	
CC	the predominant HPV subtype found in biopsies of condyloma acuminata	
CC	(benign lesions of the respiratory and genital mucosa). It was	
CC	isolated from DNA extracted from a vulva condyloma acuminatum lesion	
CC	following screening with a probe (see T13414) complementary to the 3'	
CC	end of the HPV subtype 6b L1 gene. The DNA can be used to develop	
CC	vaccines against HPV6a, to treat conditions caused by HPV6a, for	
CC	HPV serotyping, and to produce recombinant HPV6a polypeptides.	









[illegible]

RESULT 10  
ID Q03903 standard; DNA; 415 BP.  
AC Q03909;  
DE 21-AUG-1990 (first entry)  
DT 14-AUG-1990; isolate C14.  
DE DE Papilloma-virus; consensus primer; PCR; probe; ss.  
KW Papilloma-virus; consensus primer; PCR; probe; ss.  
OS Synthetic.  
PN W05002821-A.  
PD 22-MAR-1990.  
PF 29-AUG-1989; 03747.  
PR 10-MAR-1989; US-322550.  
PR 10-MAR-1989; US-243486.  
PA (CETU) Cetus Corp.  
PI Manos MM, Wright DK, Ting Y, Broker TR.  
DR MPI; 90-116005/15.  
PT Detecting and priming human papilloma-virus - using consensus primers  
PT in polymerase chain reaction to amplify particular genomic regions.  
PS Disclosure; Page 12; 33pp; English.  
CC The sequence is markedly different than the published sequence for HPV's  
CC Discovered using 11 consensus probe hybridizing to amplified DNA .  
CC The sequence enables one to isolate the entire virus from samples  
CC containing the virus. Cervical carcinoma1, C14, is a variant of HPV45.  
CC See also Q03988-Q03949.  
SQ Sequence 415 BP; 127 A; 72 C; 74 G; 142 T;

PI	Manos MM, Wright DK, Ting Y, Broker TR.
PR	WPI; 90-116005/15.
PT	Detecting and priming human papilloma-virus - using consensus primers
PT	in polymerase chain reaction to amplify particular genomic regions.
PS	Disclosure; Page 12; 33pp; English.
CC	The sequence is markedly different than the published sequence for HPV's
CC	Discovered using I1 consensus probe hybridising to amplified DNA .
CC	The sequence enables one to isolate the entire virus from samples
CC	containing the virus. Cervical carcinoma, C14, is a variant Of HPV45.
CC	See also Q03898-Q03949.
SQ	Sequence 415 BP; 127 A; 72 C; 74 G; 142 T;
<hr/>	
Query Match 18.0%; Score 274; DB 1; Length 415;	
Best Local Similarity 83.1%; Pred. No. 1.22e-176;	
Matches 344; Conservative 0; Mismatches 70; Indels 0; Gaps 0;	
<hr/>	
Db	1 tattttgttgccataatcaagttggttattctactgtagtggacaccacccgcagtaactattt 60 
OY	966 TATCTGCTGGCATTAAATTAATTTTGTACTGTGGTAGATACCACTCGTAGTACCATAATT 1025 
Db	61aacattatgtcgctctcacacaaatcctgtgccaaatcatgatcctcactaaagttaa 120 



QY 1306 GCTGAAAATAGGATCCCTATGATTAAGTTTGGAAATGGTGAATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwccvvtgkagc 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTTGGACG 1400

## RESULT 13

ID Q56374 standard; DNA; 455 BP.  
 AC Q56374;  
 DT 29-JUL-1994 (first entry)  
 DE HPV isolate 36A LI amplified region.  
 KW Human papilloma virus; amplification; polymerase chain reaction;  
 KW PCR; detection; assay; ss.  
 OS Human papilloma virus isolate 36A.  
 PN US283171-A.  
 PD 01-FEB-1994.  
 PF 29-AUG-1989; 651356.  
 PR 09-SEP-1988; US-243486.  
 PR 10-MAR-1989; US-322550.  
 PR 29-AUG-1989; WO-U03747.  
 PR 15-FEB-1991; US-651356.  
 PA (HOFF) HOFFMANN LA ROCHE INC.  
 PA (UIRP) UNIV ROCHESTER.  
 PI Broker TR, Manos MM, Ting Y, Wolinsky SM, Wright DK.  
 DR WPI; 94-04802/06.  
 DT Detection of genital human papilloma virus - by PCR amplification  
 PT using defined consensus primer pairs  
 PS Disclosure; Page 7; 13pp; English.  
 CC The sequence is that of the LI amplified region of HPV isolate 36A.  
 SQ Sequence 455 BP; 131 A; 88 C; 79 G; 144 T;

Query Match 16.5%; Score 251; DB 10; Length 455;

Best Local Similarity 74.9%; Pred. No. 1.41e-159; Mismatches 12; Indels 0; Gaps 0;

Db 1 gmcaggggwcatayaatggtatattgtggcacaatacaattgttttaacagtttagat 60  
 QY 946 GCACAGGGTCATAACAAATGCTATCTGCGCATAAATTAATTTGTTACTGTGGTAGAT 1005  
 Db 61 actactgcagcaccatctctgtgtgtgtcttactactctctctctctctctctct 120  
 QY 1006 ACCACTCGTAGTACCAATTAACAATATGCTTCTACACAGTCTCTGTACCTGGGCAA 1065  
 Db 121 tacacacctaccagttttaaagaatgccagacatggrggaatttgatggcagttt 180  
 QY 1066 TATGATGCTACCAATTAAGCAGATAGCAGACATGTTGAAGAATATGATTGCGAGTTT 1125  
 Db 181 atayttcaactgtgtaaaataacwttaactacagaggaatgtcattcattcataatg 240  
 QY 1126 ATTTTCAGTTATGACTATTAACAGATGTTGCTTACCTGCTTATTCATATGATG 1185  
 Db 241 aataccactatttggaggattggaattttgtrttacacacctctctactgtarttta 300  
 QY 1186 AATAGCAGTATTTAGAGGATTGGAATTTGGTGTCCCGCCCGCCCAACTACTAGTTG 1245  
 Db 301 gttgacacataccgttttgcatactgtgtgtaactgtgtcaactgtcaaaaggacccg 360  
 QY 1246 GTGGATACATATCGTTTGTACAAATGTTGCTTATACCTGTCAAAAGGATGCTGCACCA 1305  
 Db 361 ccagtttaaacaggacccttatgacaaactaaagttttggactgttaactcttaaggaaag 420  
 QY 1306 GCTGAAAATAGGATCCCTATGATTAAGTTTGGAAATGGTGAATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwccvvtgkagc 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTTGGACG 1400

## RESULT 14

ID Q03903 standard; DNA; 455 BP.  
 AC Q03903;

DT 24-AUG-1990 (first entry)  
 DE LI amplified region; isolate 36A.  
 KW Papilloma-virus; consensus primer; PCR; probe; ss.  
 OS Synthetic.  
 PN WO9002821-A.  
 PD 22-MAR-1990.  
 PF 29-AUG-1989; 03747.  
 PR 10-MAR-1989; US-322550.  
 PR 10-MAR-1989; US-243486.  
 PA (CETU) Cetus Corp.  
 PI Manos MM, Wright DK, Ting Y, Broker TR.  
 DR WPI; 90-116005/15.  
 DT Detecting and priming human papilloma-virus - using consensus primers  
 PT in polymerase chain reaction to amplify particular genomic regions.  
 PS Disclosure; Page 10; 33pp; English.  
 CC The sequence is markedly different than the published sequence for HPV's.  
 CC Discovered using LI consensus probe hybridising to amplified DNA.  
 CC The sequence enables one to isolate the entire virus from samples  
 CC containing the virus.  
 CC See also Q03898-Q03949.  
 SQ Sequence 455 BP; 131 A; 88 C; 79 G; 144 T; 13 Others;

Query Match 16.5%; Score 251; DB 1; Length 455;

Best Local Similarity 74.9%; Pred. No. 1.41e-159; Mismatches 341; Conservative 12; Indels 0; Gaps 0;

Db 1 gmcaggggwcatayaatggtatattgtggcacaatacaattgttttaacagtttagat 60  
 QY 946 GCACAGGGTCATAACAAATGCTATCTGCGCATAAATTAATTTGTTACTGTGGTAGAT 1005  
 Db 61 actactgcagcaccatctctgtgtgtgtcttactactctctctctctctctctct 120  
 QY 1006 ACCACTCGTAGTACCAATTAACAATATGCTTCTACACAGTCTCTGTACCTGGGCAA 1065  
 Db 121 tacacacctaccagttttaaagaatgccagacatggrggaatttgatggcagttt 180  
 QY 1066 TATGATGCTACCAATTAAGCAGATAGCAGACATGTTGAAGAATATGATTGCGAGTTT 1125  
 Db 181 atayttcaactgtgtaaaataacwttaactacagaggaatgtcattcattcataatg 240  
 QY 1126 ATTTTCAGTTATGACTATTAACAGATGTTGCTTACCTGCTTATTCATATGATG 1185  
 Db 241 aataccactatttggaggattggaattttgtrttacacacctctctactgtarttta 300  
 QY 1186 AATAGCAGTATTTAGAGGATTGGAATTTGGTGTCCCGCCCGCCCAACTACTAGTTG 1245  
 Db 301 gttgacacataccgttttgcatactgtgtgtaactgtgtcaactgtcaaaaggacccg 360  
 QY 1246 GTGGATACATATCGTTTGTACAAATGTTGCTTATACCTGTCAAAAGGATGCTGCACCA 1305  
 Db 361 ccagtttaaacaggacccttatgacaaactaaagttttggactgttaactcttaaggaaag 420  
 QY 1306 GCTGAAAATAGGATCCCTATGATTAAGTTTGGAAATGGTGAATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwccvvtgkagc 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTTGGACG 1400

## RESULT 15

ID T03499 standard; DNA; 647 BP.  
 AC T03499;  
 DT 12-JUN-1996 (first entry)  
 DE Papilloma virus major capsid protein DNA  
 KW HP-virus 29; papilloma virus; major;  
 KW capsid protein; plasmid VS93-1; DSM 9133; diagnosis;  
 KW skin carcinomas; therapy; vaccination; ds.  
 OS Papilloma virus.  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..645  
 FT /\*tag= a  
 PN DE4415743-A1.  
 PD 09-NOV-1995.



(TM)

MPsrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database which has been backtranslated into n.a. using IUPAC symbols

```
Title:
Description:
>US-08-409-122-1
(1-1524) from US08409122.seq
```

Scoring table: TABLE bktranslate2  
Gap 30

Post-processing: Minimum Match 0%  
Listing first 45 summaries

**Statistics:** Mean 58.520; Variance 269.547; scale 0.217

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	3382	47.0	505	8	R38807	BPV1 L1.	0.00e+00
2	3456	45.4	505	8	R38808	HPV1 L1.	4.89e-296
3	3049	40.0	532	17	R83016	Recombinant papilloma	2.68e-358
4	1972	25.9	495	3	P51039	Sequence of bovine pa	5.81e-159
5	1809	23.7	508	4	P30315	Sequence encoded by t	4.96e-144
6	1739	22.8	215	16	R88267	Papilloma virus major	1.25e-137
7	821	10.8	221	16	R88273	Papilloma virus major	7.11e-55
8	808	10.6	228	16	R88275	Papilloma virus major	9.93e-54
9	797	10.5	220	16	R88289	Papilloma virus major	9.22e-53
10	780	10.2	224	16	R88271	Papilloma virus major	2.88e-51
11	762	10.0	222	16	R88272	Papilloma virus major	1.09e-49
12	756	9.9	222	16	R88268	Papilloma virus major	3.66e-49
13	706	9.3	225	16	R88270	Papilloma virus major	8.57e-45
14	684	9.0	224	16	R88274	Papilloma virus major	7.04e-43
15	653	8.6	88	6	R32037	Rabbit serum reactive	3.44e-40
16	653	8.6	88	3	R14239	Seroreactive epitope	3.44e-40
17	448	5.9	39	1	P82074	Human papilloma virus	9.28e-23





























**THIS PAGE BLANK (USPTO)**